

November 2005

Published\_Applications Nucleic Acid and Published\_Applications Amino Acid database searches now generate two sets of results each. The Published\_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published\_Applications\_New databases; older published applications make up the Published\_Applications\_Main databases.

Searches run against Nucleic Acid Published\_Applications produce two sets of results, with the extensions **.rnpbm** (Published\_Applications\_NA\_Main) and **.rnpbn** (Published\_Applications\_NA\_New).

Searches run against Amino Acid Published\_Applications produce two sets of results, with the extensions **.rapbm** (Published\_Applications\_AA\_Main) and **.rapbn** (Published\_Applications\_AA\_New).

this Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2006, 20:42:45 ; Search time 625 Seconds  
(without alignments) 5731.666 Million cell updates/sec

**Title:** US-10-017-471B-19  
**Perfect score:** 4346  
**Sequence:** 1 qtcqacgacgcgcgcgggtt.....gcctagagcgcgcgggtacc 4346

Scoring table: IDENTITY\_NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 6049916 seqs, 412136615 residues

Total number of hits satisfying chosen parameters: 12099832

Minimum DB seq length: 0

**MAY 1968**

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 entries

```

Database : Published Applications_NA_New.*
1: /cgm2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
2: /cgm2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
3: /cgm2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
4: /cgm2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
5: /cgm2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
6: /cgm2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
7: /cgm2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
8: /cgm2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
9: /cgm2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
10: /cgm2_6/ptodata/2/pubpna/US16_NEW_PUB.seq.*
11: /cgm2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
C 1	82.6	1.9	10968	8	US-11-075-185-35	Sequence 35, Appl
C 2	82.6	1.9	78869	8	US-11-075-185-1	Sequence 1, Appl
C 3	78	1.8	846	7	US-10-467-657-1341	Sequence 1341, Ap
C 4	78	1.8	5706	8	US-11-052-554A-519	Sequence 519, App
C 5	74.6	1.7	88421	8	US-11-205-109-1	Sequence 1, Appl
C 6	72.6	1.7	2406	8	US-11-052-554A-550	Sequence 550, App
C 7	72	1.7	3990	8	US-11-052-554A-520	Sequence 520, App
C 8	71.8	1.7	5706	8	US-11-052-554A-519	Sequence 519, App
C 9	71.4	1.6	88421	8	US-11-205-109-1	Sequence 1, Appl
C 10	70.8	1.6	2196	8	US-11-052-554A-539	Sequence 539, App
C 11	69.8	1.6	5679	8	US-11-075-185-36	Sequence 36, Appl
C 12	69.8	1.6	78869	8	US-11-075-185-1	Sequence 1, Appl
C 13	66.6	1.5	5679	8	US-11-075-185-36	Sequence 36, Appl
C 14	66.6	1.5	11070	8	US-11-075-185-34	Sequence 34, Appl
C 15	66	1.5	2250	8	US-11-052-554A-532	Sequence 532, App
C 16	64.6	1.5	11070	8	US-11-075-185-34	Sequence 34, Appl
C 17	64.4	1.5	989	8	US-11-124-368A-159	Sequence 159, App
C 18	64.4	1.5	1980	8	US-11-129-143-177	Sequence 177, App
C 19	64.2	1.5	4617	8	US-11-052-554A-530	Sequence 530, App
C 20	63.6	1.5	3240	8	US-11-052-554A-529	Sequence 529, App
C 21	63.4	1.5	882	8	US-11-128-061-619	Sequence 619, App
C 22	63.4	1.5	882	8	US-11-128-061-4261	Sequence 4261, Ap

C 23	63.4	1.5	882	8	US-11-128-049-619	Sequence 619, App
C 24	63.4	1.5	882	8	US-11-128-049-4261	Sequence 4261, App
C 25	62.8	1.4	958	7	US-10-773-169-118	Sequence 118, App
C 26	62.8	1.4	8651	7	US-10-432-483-48	Sequence 48, App
C 27	62.6	1.4	1082	8	US-11-128-061-685	Sequence 685, App
C 28	62.6	1.4	1082	8	US-11-128-049-685	Sequence 685, App
C 29	62.4	1.4	2515	8	US-11-136-527-2896	Sequence 2896, App
C 30	62	1.4	2568	8	US-11-052-554A-533	Sequence 533, App
C 31	61.6	1.4	10968	8	US-11-075-185-35	Sequence 35, App
C 32	61.2	1.4	2302	8	US-11-110-082-20	Sequence 20, App
C 33	61.2	1.4	10524	7	US-11-073-185-37	Sequence 37, App
C 34	61	1.4	1635	7	US-10-432-483-16	Sequence 16, App
C 35	61	1.4	8651	7	US-10-432-483-48	Sequence 48, App
C 36	60.8	1.4	1263	7	US-10-858-730-149	Sequence 149, App
C 37	60.2	1.4	2304	8	US-11-052-554A-538	Sequence 538, App
C 38	60	1.4	3116	8	US-11-088-785-3	Sequence 3, App
C 39	59.8	1.4	2514	8	US-11-052-554A-543	Sequence 543, App
C 40	59.2	1.4	3408	7	US-10-858-730-40	Sequence 40, App
C 41	57.8	1.3	2562	8	US-11-052-554A-533	Sequence 533, App
C 42	57.8	1.3	3240	8	US-11-052-554A-529	Sequence 529, App
C 43	57.8	1.3	14172	8	US-11-075-185-2	Sequence 2, App
C 44	57.6	1.3	1086	8	US-11-052-554A-553	Sequence 553, App
C 45	57.4	1.3	1731	8	US-11-052-554A-548	Sequence 548, App

## ALIGNMENTS

```

RESULT 1
US-11-075-185-35/c
; Sequence 35, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010039-03
; CURRENT APPLICATION NUMBER: US/11/075.185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35
; LENGTH: 10968
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-11-075-185-35

```

Query Match	1.9%	Score 82.6;	DB 8;	Length 10968;
Best Local Similarity	43.1%;	Pred. No. 0.00012;		
Matches 624;	Conservative 0;	Mismatches 809;	Indels 16;	Gaps 4;
Qy	3	CGACGACGGCGTTCGGGTTTCGACGCCGACGGCGTACTCGTTCGCCGGCCACCGGCGACCGGG	62	
Db	4138	CGTCCAGCGCTCGGGCCCGAGGTCGACGAGGCGCACCGCCGCGCGGGGTTCTCTCTGGC	4079	
Qy	63	TCCTGGCTCGATGACCGACCGCATCGAGGACGTCGGCTGGCGGGCTCCTGATAGTAGACGG	122	
Db	4078	GCGCCGTGCGCACGAGGCCCCACAGCGCGCGGTGGGTCAACGCCCGGACGCCCTGCTGCGG	4019	
Qy	123	CCCGCGCGGGGACGCACATCGACGTCCTCCACTCGGCCCGCCCGGAAAGTAGACAC	182	
Db	4018	GCCTCGGGCGCACAGCGCCGGGTCAACACACACAGCTCGTGTCCGCGAGCGCGCCT	3959	
Qy	183	CGCACCCGACGTCGACGCCATGGGAGGGCCAGCTCCGCGGACCGATCAACCCCTGGCTTC	242	
Db	3958	CCGAGAGCCACCCTGCAACGACGACGAGGACCCGCTCGCTCGCGCTGCGCGCCCGCCGA	3899	
Qy	243	GGCGCAAGGCTTCGGCTGTGTCGCGCCGCCAGATGCGGAAAGGGCTCGCGGGCGCGGCCG	302	

[illegible]

Db	2831	CGACCGCGCGT	GAGCA	CGGGTGCTT	ACCGCGCAGGAGCCCTCCGCTCTTCAGCCCT	2772	
Qy	1380	AAGGGCGCGGAC	CGGT	CACGAGCAGCGCGCT	CGCTGCGCGCTCCAGCAGCAGCATG	1439	
Db	2771	GCCCGCGCGCG	CGCGCT	TCGAGCAGTAGCGCT	GCCGCTGGACCGGTAGGT	2712	
Qy	1440	CCCGGTACG	1448				
Db	2711	TCGCGCACG	2703				
RESULT 2							
US-11-075-185-1/c							
; Sequence 1, Application US/11075185							
; Publication No. US20050266434A1							
; GENERAL INFORMATION:							
; APPLICANT: REEVES, CHRISTOPHER D							
; APPLICANT: JULIEN, BRYAN							
; APPLICANT: REID, RALPH							
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS							
; FILE REFERENCE: 010099.03							
; CURRENT APPLICATION NUMBER: US/11/075,185							
; CURRENT FILING DATE: 2005-03-07							
; PRIOR APPLICATION NUMBER: US 60/551,103							
; PRIOR FILING DATE: 2004-03-08							
; PRIOR APPLICATION NUMBER: US 60/568,290							
; PRIOR FILING DATE: 2004-05-04							
; NUMBER OF SEQ ID NOS: 61							
; SOFTWARE: Patentin version 3.3							
; SEQ ID NO 1							
; LENGTH: 78869							
; TYPE: DNA							
; ORGANISM: Sorangium cellulosum							
US-11-075-185-1							
Query Match 1.9%; Score 82.6; DB 8; Length 78869;							
Best Local Similarity 43.1%; Pred. No. 0.00014;							
Matches 624; Conservative 0; Mismatches 809; Indels 16; Gaps 4;							
Qy	3	CGACGCGCGT	CGGGTTCG	ACGCGCAGCGGT	ACTTCTCCGCGCCACCGGCA	62	
Db	19405	CGTCCAGCGCT	CGGGCCCGAGT	CGACGAGCGC	ACCGCCCGCGCGGGTTCTCTGC	19346	
Qy	63	TCGCGCTCAT	GACCGC	CGCATCGAG	ACGTCGCGCTGCGGCTCTGTATAGT	122	
Db	19345	GCGCGTGC	ACGACGCC	CCACAGCGCGCT	CAACGCGCAGCGCTCGT	19286	
Qy	123	CCCGCGCGG	CGACGCA	CATGCA	CGTCACTCCACATGCGGCC	182	
Db	19285	GCCCGCGG	CCACAGCGCGCGT	CACCA	CACAGCTCGCTGTCCGCGAGCCGCGCT	19226	
Qy	183	CGACCGGCA	CGTAGCGC	NTGGAGGCGC	ACGTCGCGGACGGATCACCCCTGGCTTC	242	
Db	19225	CCGAGAGCC	ACCCTCGA	CAGCGA	CAGGACCGCTGCGCGCGCGCGCA	19166	
Qy	243	GGCGAAGG	CTCCGGTG	TGTCGCGCGCCAGAT	CGGAA	CGGCTGCGGCGCGCGCG	302
Db	19165	CCAGGCGCT	CGCGGGT	CGCCGCGCGCCG	GTGATCACAGCGCGCGCGCGCGCG	19106	
Qy	303	TCACTTCCG	CA	CGGTCGCGCGT	TTGGCCCGCGCGCGCTG	CGCGTCCGGCGC	362
Db	19105	CGCCCTCCG	CA	CGCGCGATG	ATCGCGGACGCGCGCTG	CGCGCGCGCGCGCGCG	19046
Qy	363	TCTCCAGAA	CGGT	CGGCA	GGGCGCTCGGAT	CTTCGAGAGCATCTGCTCGCGCGCG	422
Db	19045	CCGCGAGCG	CCCGT	CGCGCG	CGAGCA	CGATGTGCTCTCCGCGCGCCACGC	18989
Qy	423	CCTCCAGCC	CCAGCT	CTCTCGAA	ACGAGTGA	CCGCCACGACCTCTCGAGAGCCCGCGCG	482
Db	18988	CTTCCAGCG	CA	CGTCCG	TCA	CGGTACAGGTACAGTCTGCTCGCACCGCGCG	18929
Qy	483	TCGTCTG	CGTGGCGG	TGATCG	CGTCCGGGCGCGGA	CAGCGCGTACGCCCGTCTCTGTTGC	542



Db 18928 CAGCTTTCCGACGCTGCTCGCGCTCGCCACCGCGCCCGGAAAGTCCCTCGCGCGCGA 18869  
Qy 543 CGGGCCGGGTGAGTCCCTGACCGCCTTTCGCGGATGTCGCGGATGCGACGAGCGAA 602  
Db 18868 CGGGTCCCTGCGCTCGCGCTGATCGACGCGCGAGCTGCTCCCATCGCTGG 18809  
Qy 603 CGGGGAGCTGCTGACGCGCGGAGACACCGCTCGGACCGGATGCGGCGCGCCAG- 661  
Db 18808 GGTCTCGACCTTCCACGCGCGCGCGAGCGCGCTGCCCCCGCGCTCGAGCGCCAGT 18749  
Qy 662 --GACAGAGCTGTTGACATGAGGTCTTGGCCCGCAGGAAGTCCAGTCTAGCCGGACT 719  
Db 18748 CCGACACGCGAGGGCATCGGATGCTCTCCGCGGACGAGCTCGCGCGAGCTCG 18689  
Qy 720 CGCGTACGCGCGCTGCTTCTCGCGCTCGCGCGCTGATGAAAGTCTCGCGCGCGCTT 779  
Db 18688 CCGCAGGAGGCGCTGGAGCGCGCATCGAGCAAGCGCGATGCAAGCGGAAGTCTCGC 18629  
Qy 780 CCCCACCGCGAGGATGACAGCTTACAGAGTCCGAGCGCGGCTCGCGCGCGCG 839  
Db 18628 CCGCTCGGAGACCCCTCGGGCAACGCGAGGTGGCGGTAGATCGACCGCCCGCC 18569  
Qy 840 CGCGAAGCGCTGCTCGCTCGCTCGCTCGCGCTGATGAAAGTCTCGCGCGCGCTCGCA 899  
Db 18568 AGCGCGCGCTCAGCGCCCGGAAAGCGCGGCGGACCGCTCGAGCGCTCGACCCAC 18509  
Qy 900 CCCCCCTTGAAGGCGCGCTCGAGCGCGCGCTGCTCGCGCGCGCGAGACTCGA 959  
Db 18508 GGTGACGTGACCGCTCGGCTCGGCCCC--GCGCGCGCGCACGCGCGCGCGCCACGC 18454  
Qy 960 GGTTCGCGCGGTGACGCTGACCCGTTTCGCGCGCGCGCGCGAGACTCTGACGCGGACCG 1019  
Db 18453 GGTTCGCGCGCGGTGCGCGCGCGCTGCGCGCGCGCGCGCGCGCTGCGCGCTCGACGCGCG 18394  
Qy 1020 TCCGCGCGCGAGTGGCGCACTGACGCGCGCGCACACCGCTGACGCGCGCTCAAGAA 1079  
Db 18393 GTCTGCGCGCGCTGCGCGCGCGCTGTGACCGCGAGCGCGCGCGCGCGCGCTCGCTCCAA 18334  
Qy 1080 TCACCTCGCGCGCTCTCTCGCGCGAGCGCGCGCGCTCGCGCGCGCTCAATATGAGGG 1139  
Db 18333 CGCTTCCACCTGACCTGACGCGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCTC 18274  
Qy 1140 AAGGCGAGGATGCTCCCGCGCGCGCGCAACCGCGGATGTTGCGCGCGCGCGCGCGCTT 1199  
Db 18273 GATGCTGAGCTCTGACACCGAGGCGAGTCCGACGCGATCGCGCGCGCGCGCGCGCT- 18215  
Qy 1200 CAGCGCGAGAACCGCGCGCGCGACAGCGGTGAGGAGACGAGGCTGCGCTCTGATGC 1259  
Db 18214 ----CGACGAACCGCTGCGCGCGGAGATCACCGTGCAGCACCAAGTGGTCCAGGACC 18159  
Qy 1260 CGGCTGACCGCGACGCTGTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGGATC 1319  
Db 18158 CAGCGTGTCTGACCGTTCAGACGCGTTCGCTCAGGACGCTAGCGCTCGCGGTGCGCCAGC 18099  
Qy 1320 CAGCACGCGGCTGCGAACTTCGCGTACCGGACGAAACCGGCTGCGCGCGATCGACGCGCAG 1379  
Db 18098 CGCACCGCGCGCTGAGCCACCGGCTGCTTACCGCGCAGGAGCGCTCGCTCTCCAGCCCT 18039  
Qy 1380 AAGGGCGCGGACCGGTTCAGAGGACGCGCGCTGCGCTGCGCGCTCGAGGAGGACGATG 1439  
Db 18038 GCGCGCGCGCGCGCGCTTCGAGCCAGTAGCGCTGCGCTGGAACGCGTAGTTCGCGCAAG 17979  
Qy 1440 CCGGTACG 1448  
Db 17978 TCCGCGACG 17970

RESULT 3  
US-10-467-657-1341/c  
; Sequence 1341, Application US/10467657  
; Publication No. US20050260581A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA

; APPLICANT: FONTANA Maria Rita  
; APPLICANT: PIZZA Mariagrazia  
; APPLICANT: MASIGNANI Vega  
; APPLICANT: MONACI Elisabetta  
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/467,657  
; CURRENT FILING DATE: 2003-08-11  
; PRIOR APPLICATION NUMBER: GB-0103424.8  
; PRIOR FILING DATE: 2001-02-12  
; NUMBER OF SEQ ID NOS: 9218  
; SOFTWARE: SeqWin99, version 1.04  
; SEQ ID NO 1341  
; LENGTH: 846  
; TYPE: DNA  
; ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-1341

Query Match 1.8%; Score 78; DB 7; Length 846;  
Best Local Similarity 50.4%; Pred. No. 0.00055;  
Matches 251; Conservative 0; Mismatches 235; Indels 12; Gaps 2;

Qy 3031 ATCGTTCCGCTGTGGCATCGACGTACTGGCGGTGATCCACCGTGAATCGTCGAGGCC 3090  
Db 839 AGCATGCCCGCTTGACGTGCGCGTTTGACCGGTAATGTTTTCGCTGATCGGAAGCG 780  
Qy 3091 AGAAGGCGCACACGTCGCGGATGTCGCGGTCTCGCGATGCGGTTGAACACGAGTTG 3150  
Db 779 AGGAACAGAACCGCATCGCGATGCTTTCGCGGTGCGCGAATCTGCCAAGCGGTTTG 720  
Qy 3151 GCGGCCAGTCCGCGCGCGCTCGGGGTCTGCGCGCGGTTCATGTCCTCTCC 3210  
Db 719 GCGGTAAATGTTTGGCGGTTCCTTCGGGCGAGGCGCG--GTCATATCGGTATCG 666  
Qy 3211 ACGAACCCCGCGCGCACCGCTTGACCGTGTATCCCGCTTCCCGAGTTGCCCTGCCAGG 3270  
Db 665 ATAAAGCCCGGGCGGACGAGTTGACGCTTAATGCCCGCTCGCGACTTCGCGGCCATA 606  
Qy 3271 GCGAGCTGAGCGGTGTCACCGCACCTTGGTTCATCGGCTATCCGATGAGATCGGGGAAC 3330  
Db 605 GATTTGGCAAAACCATCAAGCCCGCTTTTCGGCGGCAATATGTTGTTGCCCGCATTTG 546  
Qy 3331 GCGCGCGCGGTGCGGGCAGAGATGTTGATGATCCGCGCGCG--TCGCGCGAGT 3384  
Db 545 CCATCACCGCGACGACGAGGTGATGTTGATGATGCGCGCGCGCTGCTTCATCATG 486  
Qy 3385 GCTTTTCAGTTCGCTGCTGACAGGAAACAGCGGTGCCCGAGCTTCAACGCGACCGTCCG 3444  
Db 485 CGCGCAAGACGGCTTTAGAGCGCGGAACACGGATTGAGGTTGACCTGCTGATGTCG 426  
Qy 3445 TCGAAGACTCTCTCGGTGACTTCGCTGATTCGCTCGCGAGCGCTGACGCCCGTTGTC 3504  
Db 425 TCCCACTCTCTCTCTTCATAGCATCAACAGTTGTCGCGGTGATGCGCGGTGTTG 366  
Qy 3505 ACCAGGATGCGAACTCG 3522  
Db 365 ACCAGAATGTCAGTTTG 348

RESULT 4  
US-11-052-554A-519/c  
; Sequence 519, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

```

; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 519
; LENGTH: 5706
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-519

Query Match      1.8%; Score 78; DB 8; Length 5706;
Best Local Similarity 43.8%; Pred. No. 0.00062;
Matches 577; Conservative 0; Mismatches 725; Indels 14; Gaps 5;

Qy 3038 GCGCTGTGGCATCAACGTACTGGCGCGGTGATCCACCGTGTGTCGGAGGCCAGAAAGG 3097
Db 5188 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5129

Qy 3098 CCACCAAGTCGGCGCATGTGTCGCGGGTCTGCCGATGCGG--TTGAACACGAGTTGGCGGC 3155
Db 5128 GCGCGAATCCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 5069

Qy 3156 CAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3215
Db 5068 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5009

Qy 3216 ACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3275
Db 5008 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4949

Qy 3276 CGTGAGCGGTGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3335
Db 4948 CGACCCCGCGCTTGGCGG--CTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4891

Qy 3336 CCGCGGTCCGCGCGACACGAGATGTTGATGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3395
Db 4890 GCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 4831

Qy 3396 GTGCTGACACGAGAACAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3455
Db 4830 GCGCGCGCGGTTCGCGCGGTGTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4771

Qy 3456 CTCGCTGACTTCCTGATCCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3515
Db 4770 GGTGCGCGCGACCGCGCGATAGCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4711

Qy 3516 GAATCTGCGCGCGCACTCCGAATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3575
Db 4710 GCGCGCGCGCGGTGCGCGCGTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4651

Qy 3576 GCGGTACCCACGAGCGCGAGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3635
Db 4650 GCCTTTGCGCGCTTGGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4591

Qy 3636 GATGCTCTCAGCGGTCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3695
Db 4590 GTCCCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4531

Qy 3696 GCGCGCGGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3751
Db 4530 GTCGCGCGGAACACCGCGCGCGAATCCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCG 4471

Qy 3752 -CGGTACACGAGCGGTCTTGCCTTCAGCGGTCTTCCATACCTCGTCCCATGTGCGAGCG 3810
Db 4470 GCGCGCGTACACCGGTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4411

Qy 3811 ATATCAGCGCGCGCGGTGCGTGAGCGACCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3870
Db 4410 CTGCGCGTCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4351

Qy 3871 GTCAAGCGCTACTGTGTACCGCGTACAGCGGCGCGGAGTGGCGCGGTGTGAACGCGGTGG 3930
Db 4350 GCGCGCGCGC--GTTGCGCGCGACCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGGT 4293

```

```

Qy 3931 GCAGATCGCGCGCGCGCGCACCGGGAACCGCGCGCGGTCAAGGGGTCAAGGGGTCAAGGGG 3990
Db 4292 CCGTCTTGGAGAGGGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 4233

Qy 3991 CCGCGCAGGCGCGTCAAGGACACGACCGGATCAAGGTCAAGGTCAAGGTCAAGGTCAAGGT 4050
Db 4232 CCGCGCATACCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGT 4173

Qy 4051 TCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4110
Db 4172 TCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4113

Qy 4111 TGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4170
Db 4112 TTGCG---GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4056

Qy 4171 GCGCGCAGCATCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4230
Db 4055 AGCGCGGTGTGTTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3996

Qy 4231 TCAGGACGTCTAGACCGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4290
Db 3995 TGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936

Qy 4291 TCAGCGCGGTGCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4346
Db 3935 CCGGTGCGCGCGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3880

RESULT 5
US-11-205-109-1/c
; Sequence 1, Application US/11205109
; Publication NO. US20050287641A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
; FILE REFERENCE: 3002-2US
; CURRENT APPLICATION NUMBER: US/11/205,109
; PRIOR FILING DATE: 2005-08-17
; PRIOR APPLICATION NUMBER: US/09/976,059
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/239,924
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 88421
; TYPE: DNA
; ORGANISM: Actinoplanes sp.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2077)..(3078)
; OTHER INFORMATION: ORF 1; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3118)..(4032)
; OTHER INFORMATION: ORF 2; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4038)..(5048)
; OTHER INFORMATION: ORF 3; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6665)..(5814)
; OTHER INFORMATION: ORF 4; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7703)..(6693)
; OTHER INFORMATION: ORF 5; negative strandedness
; FEATURE:
; NAME/KEY: misc feature

```

LOCATION: (9464)..(8130)  
OTHER INFORMATION: ORF 6; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (9691)..(10761)  
OTHER INFORMATION: ORF 7; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (12751)..(10829)  
OTHER INFORMATION: ORF 8; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (13617)..(12802)  
OTHER INFORMATION: ORF 9; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (15203)..(13614)  
OTHER INFORMATION: ORF 10; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (15591)..(15863)  
OTHER INFORMATION: ORF 11; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (15880)..(19035)  
OTHER INFORMATION: ORF 12; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (19032)..(39713)  
OTHER INFORMATION: ORF 13; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (39713)..(65800)  
OTHER INFORMATION: ORF 14; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (65826)..(66530)  
OTHER INFORMATION: ORF 15; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (66546)..(67370)  
OTHER INFORMATION: ORF 16; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (67384)..(70059)  
OTHER INFORMATION: ORF 17; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (70099)..(70662)  
OTHER INFORMATION: ORF 18; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (70659)..(71906)  
OTHER INFORMATION: ORF 19; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (73439)..(71964)  
OTHER INFORMATION: ORF 20; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (74216)..(73563)  
OTHER INFORMATION: ORF 21; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (75424)..(74213)  
OTHER INFORMATION: ORF 22; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (75535)..(76464)  
OTHER INFORMATION: ORF 23; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (78110)..(76449)

OTHER INFORMATION: ORF 24; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (79864)..(78107)  
OTHER INFORMATION: ORF 25; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (81624)..(79861)  
OTHER INFORMATION: ORF 26; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (81909)..(81682)  
OTHER INFORMATION: ORF 27; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (82346)..(82062)  
OTHER INFORMATION: ORF 28; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (82587)..(84446)  
OTHER INFORMATION: ORF 29; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (84481)..(85548)  
OTHER INFORMATION: ORF 30; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (85556)..(86845)  
OTHER INFORMATION: ORF 31; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (87372)..(86803)  
OTHER INFORMATION: ORF 32; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (87494)..(88420)  
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only  
US-11-205-109-1

Query Match 1.7%; Score 74.6; DB 8; Length 88421;  
Best Local Similarity 48.7%; Pred. No. 0.0025;  
Matches 320; Conservative 0; Mismatches 304; Indels 33; Gaps 3;  
QY 3152 CGGCCAGTCCGCGCCCTCGGGGGTCTGCGCGCGCGGTTCATGTCCGCTCCA 3211  
Db 67206 CQTCGAGCGACCGGCGGATCTCCGACAGCATCCGCCCTCGAACTCGCGCA 67147  
QY 3212 CGAAACCGCGCCACCGCTTGACCGTGATCCCGCTTCCCGCAGTTGCCCTGGCCAGG 3271  
Db 67146 CCGAAACCGCGGACGCGGTTGACCCGACCTTCGCTGGCGCGTAGTCGAGCGGCGG 67087  
QY 3272 CGAGCGTGAGCGTGTCCACCGCACCCCTTGGTTCATCGCGATCCGATGGAATCGGGAAACG 3331  
Db 67086 CCGCGGTGAGCCCGACGATCGCGTCTTGGCGGACGATAGCCGCGAAGTCCCGTAGC 67027  
QY 3332 CGCGCGGGTCCGCGCAGACGAGATGTTGATGATCCGCCCGCGCGCTCGCGCATGCTTTCA 3391  
Db 67026 CGACCTGACCGCGCGTCCGAGCGATGTTGACGATCTGCGGAGCGGCTCGGTCACT 66967  
QY 3392 GTCCGTGTGGAC-----CAGGAACAGCGGTGCGCGGACGTTGACGGCGACAGTCGCT 3445  
Db 66966 CCGCGCGACCGCGCCCGCTCATGCGCCACGCGCCCGAGAGGTCGACGTCACTCATGCT 66907  
QY 3446 CGAAGACCTCTCGGTGACTTCCGTGATCCGTCCGAGCGCTGACGCCCGCGGTGTTCA 3505  
Db 66906 GCCACTCGTCTCGGTGATCTGTGGTGA---CTTTCCGAGCGCGCGCGATGCCCG 66850  
QY 3506 CCAGGATGCGAACTCGGCGGCACTCCGAACTCGCCCATCCCGGCTCGAAGCGCGCT 3565  
Db 66849 CGTTGTTGACGAGCAGTCGATCGCGCGCGAAGCGGTGATCGCCCGATCGGCGCGCG 66790  
QY 3566 AGAGCGCGCGCTGATCCACGACCGCGAGTTCGGCCCGGATGCCCAACGCTGTCGCG 3625



;; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
;; FILE REFERENCE: 30853/40359A  
;; CURRENT APPLICATION NUMBER: US/11/052,554A  
;; CURRENT FILING DATE: 2005-02-07  
;; PRIOR APPLICATION NUMBER: US 60/589,227  
;; PRIOR FILING DATE: 2004-07-20  
;; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
;; PRIOR FILING DATE: 2004-02-06  
;; NUMBER OF SEQ ID NOS: 763  
;; SOFTWARE: PatentIn version 3.3  
;; SEQ ID NO 520  
;; LENGTH: 3990  
;; TYPE: DNA  
;; ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-520

Query Match 1.7%; Score 72; DB 8; Length 3990;  
Best Local Similarity 42.7%; Pred. No. 0.0052;  
Matches 369; Conservative 0; Mismatches 495; Indels 0; Gaps 0;  
QY 3483 GCCGCTGACGCCCGCGTGTGTTCCACGAGGATGTCGAACCTGGCGCGCACTCCGAACTCGCC 3542  
DB 1893 GCCGCCCCCGCGCCCGCGCGGACGTGGGTTGCGCGCGTTCGCGCGCGTTCGCGCC 1834  
QY 3543 CATCCCGCGCTCGAACCGCGCTAGAGCGCGCGCGTCAACCAAGAGCCCGAGTTCGCGC 3602  
DB 1833 GTTGC CGCGCTGTTGCGGGTGGCGCGCGCGCACCGTCTGCACCGCGGTGATCCCGC 1774  
QY 3603 CGGATGCGCAACGCTGTCCCGCGCTGCTCCGATGGTCTCGAGGTTCTTCGCGCGCGC 3662  
DB 1773 GCCCGCATCCACAGACACCGATGCCGCGGTACCGCGTGTGCGCGCGTTCGCCACC 1714  
QY 3663 CCCTCGCTGCTGCTAGTGAAGTCCACAGAGCGCCCGTCCGCGCGCGAGCGCGCGAGGCG 3722  
DB 1713 AACTCATCGCGCGTATCGAAGTCCCTTGGACCGCGTTCGCCCATCACGCCCAT 1654  
QY 3723 GATACCGGCTCGATGCCCGCGTTCGCCCGGTCAACAGGCGGTGTTGCCCTTCAGCGG 3782  
DB 1653 GCCGCGCGCGCGCGTTCGCCCGCGCGCGCGCGCACCATGCTGCGCGCTCTGCGTGGT 1594  
QY 3783 TCTTCATACCTGCTCCATGTCAGGATATCAGCGCGCGCGCGTGTGAGCGACCCAT 3842  
DB 1593 GCGTCAAGCGCTTACCGCGCTTGCACCGCTCCACCGCGCGCGCGCGCGCGCGTTC 1534  
QY 3843 GCGCGCGCTGCGCGCTTCAATCGAGCGTCAAGCTTACCGTCAACCGCGTCAAGCGG 3902  
DB 1533 GCGCGCTTGC CGCGCTGCGTGCATTCGCGCTGCGCGCGCGCGCGCGCGCGCGTTC 1474  
QY 3903 GCGGAGTGGCGCGGTGGACGCGCTGGGCGCGAGATGGGCGCGCGCGCGCGCGCGCG 3962  
DB 1473 GCGGTGGCGCGTTCGCGCGCTTGC CGCGCTTACCGCGCGCGCGCGCGCGCGCGTTC 1414  
QY 3963 GCGGTGAGGCTCAGGCTCAGGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4022  
DB 1413 GCGTGTGGGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1354  
QY 4023 GAGGTGCGCGTGCACAGCGCGCACAGTCTCTCGCGCGCGCGCGCGCGCGCGCGTTC 4082  
DB 1353 GCGGTAGGCTGCGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTC 1294  
QY 4083 CCAGAGCGCGGTCCGCGCGAGTCTGAGTGGACGCTGGCGCGCGCGCGCGCGCGCGGT 4142  
DB 1293 GCGCGCGCATTTGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTC 1234  
QY 4143 GCGCGAGCGCGAGAGCGCGGAGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTC 4202  
DB 1233 GCGCGCGATGAGAGTGGCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTC 1174  
QY 4203 AGCGCGCGCGCGGTTCAGGCTCTCTCAGGAGCTCTAGACCGCGCGCGCGCGCGCG 4262  
DB 1173 GCTGGCG 1114  
QY 4263 CCGCGAGCGAG 4322

DB 1113 CCCGCGCGCGCGCGGTTCACCGCGCTTATCCCGAGCAATCAAGGACGAGAAATCC 1054  
QY 4323 CGACGCGTAGAGCGCGCGCGGTAC 4346  
DB 1053 GGCGCGCGCGCGCGCGCGGTCCC 1030  
RESULT 8  
US-11-052-554A-519  
;; Sequence 519, Application US/11052554A  
;; Publication No. US2005028866A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Sachdeva, et al.  
;; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
;; FILE REFERENCE: 30853/40359A  
;; CURRENT APPLICATION NUMBER: US/11/052,554A  
;; CURRENT FILING DATE: 2005-02-07  
;; PRIOR APPLICATION NUMBER: US 60/589,227  
;; PRIOR FILING DATE: 2004-07-20  
;; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
;; PRIOR FILING DATE: 2004-02-06  
;; NUMBER OF SEQ ID NOS: 763  
;; SOFTWARE: PatentIn version 3.3  
;; SEQ ID NO 519  
;; LENGTH: 5706  
;; TYPE: DNA  
;; ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-519

Query Match 1.7%; Score 71.8; DB 8; Length 5706;  
Best Local Similarity 42.2%; Pred. No. 0.0057;  
Matches 606; Conservative 0; Mismatches 817; Indels 14; Gaps 3;  
QY 2909 CGCGGAAGCGCGCGCGCGCGTACCGACCGCGCGTCCCGAGCGCGCGCGCGCGCGCG 2968  
DB 3453 CGCGCGCAAGCGCGCGCGCGTGTGCGGGAACCTTCGGCGGTTCGCGCGCGCAC 3512  
QY 2969 CTACGCGCGCGCGCGCGCGCGTGTAGGTCTGCGCTGCGTACCGAAGCGTGGCGGGTCA 3028  
DB 3513 CGCGCGGAGCGCGCGCGCGTGTGCGCGCGCGTTCGCGCGCGCGCGCGCG 3572  
QY 3029 GAATCGTTCGCGCTGTGCGATCGACGTACTGCGCGCGTGTATCCACCGTGTGCGGAG 3088  
DB 3573 TAAGCGCGCGCGCGCGGAAACCGCGGTGTTGGCGGTGACCGCGCGCGAGCGCGCG 3632  
QY 3089 CCAGAAAGCGCACACGTGCGCGATGTCGCGGTCTGCGGATGCGTGTGAACAGGAGT 3148  
DB 3633 TCTCGCGCTGCGCGCTCTCCGCGCTTTGACGCGCGCGCAAGCGCGCGCGCGCGCG 3692  
QY 3149 TGGCGCGCAGTCCCGCGCGCGCTTCGCGCGGTCTGCGCGCGCGTTCGTTTCATGTCG 3208  
DB 3693 CAGCGCGCGCGCGCGCGCGTCAACCGCGG-----CCGCGCGCGCGCGCGCGCG 3745  
QY 3209 CCAGAAACCGCGCGCGCGCGTGTGACCGTGTATCCCGCGTTCGCCAGTTCGCGCGCA 3268  
DB 3746 GGGCGCGGTGACGCGCGCGCGCGCGCGTGTGACCGCGCGCGCGCGCGCGCGTTCGCG 3805  
QY 3269 GGGCGAGGTGAGC-GTGTCCACCGCACCTTGGTTCATCGGTATCCGATGCGATCGCGG 3327  
DB 3806 GCCAAGCGCGCATCGCGCGCGCGCGCGCAACCGCGGATTCGCGCGCGGTTCGCGCGG 3865  
QY 3328 AACCGCGCGCGGTTCGCGCGAGACGAGATGTTGATGATCCGCGCGCGTTCGCGCGAGT 3387  
DB 3866 ACGCGCGGAGACCGCGCGCAACCGCGGATTCGCGCGCGGTTCGCGCGCGCGCGGATCG 3925  
QY 3388 TTCACTCGTGTGAGACCGAGAACCGGTGCCCGGACGTTCACGCGCGACAGTTCGCGT 3447  
DB 3926 GCGGCACCG 3985  
QY 3448 AAGACCTCTCGTGTGATTCGCGGTGATTCGCGGTCCGAGCGCGCTGACGCGCGGTGTT 3507

Db	3986	ACGCGCGCCAAAGTGTGTGTCGGCGGGCCACGCGGGCCAAAGCGGGCTGAAACA	4045
Qy	3508	AGGATGTGCAACTCGGCGCGCACTCCGGAACCTCGGCCATCCCGCGGTGCGAACGCGCGGTAG	3567
Db	4046	GCACCGGGCTAGCCAGCGCCGCCAGCGGTGACGCGGCAACAGCGGGCGCGCGGGCCG	4105
Qy	3568	AGCGGGCGGGGTCAACCAAGACGCGCAGGTTTCGGCCCGGATGGCCAAACGCTGTCCGCCG	3627
Db	4106	CGGCGAACGCGGGCGACCGGCGACGGCTTTATTCGGCGGGTCCCGCGGCACCGCGGGACCG	4165
Qy	3628	CTGCTCCGGATGGTCTCGACGGTCTCTCGCGCGCGCCGCTCGTCTGCTGCGGTAGTGGACT	3687
Db	4166	CGGCGACGCGGGGTTCGGGGCTGCGGGCTGCGCAACCGCGGAAACGCGGGCAACGCCGGA	4225
Qy	3688	GCCACGACGCCCGTCCGCGGCGACGCCGAGGGCGATACCGGTCGATGCCCCCGGCTT	3747
Db	4226	TGCGCGGGCGCGGGCGCGCGGGCGGACGCGGGGGCGGCGACGCGGGCGCTCTCCC	4285
Qy	3748	CCCCCGGTCAACGAGGGCGGTTTGCCCTTCAGCGGTTCTTCATACCTCGTCCCATGTGCA	3807
Db	4286	AAGACGGCAACGGGTTTCGCGCGGGGCCAAAGCGGCCCAAGCGGGTTCGGCGGCAACGCCG	4345
Qy	3808	CGCATATCAGCCCCCGCGTGTGCGGTGAGCGACCCATGCGCGCGCTCGCGCCGTTTCGAATCG	3867
Db	4346	CGCGCGGGCATCAACGGGGCCGCGCGGCACCGCGGCACCGCGGGCGCGGTGGTGACG	4405
Qy	3868	ACGGTCAACGCTTACTGTGACCGGTTCAGACGGGGCGGAGTGGCCCGGTGTGACGGCT	3927
Db	4406	GCCAGAACGGAAACACAGAGGCTGGCGAGCGAGGGCGCGCGCGGCAAGGGCGGTGACG	4465
Qy	3928	GGGGCCAGATCGGGCGCGGCACCGGGGAAACCGCGCGCGTCAAGGGTCAAGGGTCCCG	3987
Db	4466	CGGCGCAAGCGGGCATCGCGGGGCCCGCGGCAACGCGGATTCGGCGCGCGGTGTCCCG	4525
Qy	3988	GGACCGCCAGCGCGGTGAG-----GGCAACCGACCGGATCGAGGTTCGGGCGTGCACACG	4041
Db	4526	GGACGCGGGATCGCGGGCACCGCGGGGCGGGGGCGCGCGCGCGCGCGCGCGCGACG	4585
Qy	4042	GGCCACAGTCTTCGCGGCCAGTCTCGACTCTGATCGGATACAGAGCCCGGTCCGGCGCG	4101
Db	4586	GGGACCCAGATTGACCGCGGGCCAAAGTGGTGTCCCGCGGCACGCGGGGCCAAGCGGCGCA	4645
Qy	4102	AGTCTGAGCTGAGCTGCGCGCGGGTGGGTGAGGGCGGTTGCGCCAGGGGCGGAAGGGCG	4161
Db	4646	AAGCGGCTTGAAACAGACACGGGGTAGCCAGCGCGCTAGCGGTGACGGCGCAACGCGCG	4705
Qy	4162	GGGAGGTTCGGCGGAGGATCATGTGGGCGGGCGGGTTCGAAAACGCGCGCGCGGGGTCC	4221
Db	4706	GGGCGCGGGGCGCGGGCCAAACGGCGCGGACCGGCTTTATTCGGCGGGTTCGGGCG	4765
Qy	4222	CAGGGCTCTCTCAGAGCTGTAGACCCGCGCAACCGCGCTTCGCCACGAGGGCGGACGGCC	4281
Db	4766	GCACCGGGCGGACCGCGGGCGACCGCGGGCTTCGCGGGCTTGGCCAAACACCGCGGAACCG	4825
Qy	4282	CGCGCCAGCTCCGCGGTGTGCGTTCGCGCGCGGCGGAGCGACGACGCTTAGAGCGCG	4338
Db	4826	CGGGCAACGCGCGGTATCGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	4882

## RESULT 9

```

RESUMI 3
US-11-205-109-1
; Sequence 1, Application US/11205109
; Publication NO. US20050287641A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staiffa, Alfredo
; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
; FILE REFERENCE: 3002-2US
; CURRENT APPLICATION NUMBER: US/11/205,109
; CURRENT FILING DATE: 2005-08-17
; PRIOR APPLICATION NUMBER: US/09/976,059
; PRIOR FILING DATE: 2001-10-15

```





Db 40915 ACGCCCGCGGTCTCAACCGGGGCACTGATCGGGCGCGCGCTTTCGAGCGCGGTC 40974  
Qy 3784 CTTTCATACCTCGTCCCATGTGACGATATACAGCCCGCGGTGCGTGACGACCACTG 3843  
Db 40975 GGTGAGCGGTGCGCGCGTCTGCGCCCGCGCTCGAGGCGTCTGCGCGCGCGCGCG 41034  
Qy 3844 GCGCGCGTCCGGCGTTTGAATCGAGCGTCAAGCGTACAGCTACCTGTGACCGCGTCAGACGSGG 3903  
Db 41035 CACGCGCGCGCGGACCTCGACCTGCTCTCCCGCGGACCGCGCGTGAATCTTGGCGG 41094  
Qy 3904 CCGAGTTCGCGCGTTCGACCGGCTGGGCGGACAGTCCGCGCGCGCGCACGGGGAACCGGCG 3963  
Db 41095 CTGGAACACACCGCGGCTCCGGCGCGCGCGGACTGGTGGCGGACCTG----TTGCGG 41150  
Qy 3964 CCGGTTCAGGGTTCAGGGTTCGCGGGAACCGCGGACCGCGGTCGAGGGAACCGGAGTCG 4023  
Db 41151 CCGAGCGCGCGCGCACCGCGGACCGGTCGCGGCTCGCGGCGCGCGGACCGGAGCTGACCT 41210  
Qy 4024 AGTTCGGCGGTCCACGCGGCGCACCGAGTCTCGCGGCGCGGCTCGGAGCTGCTAGCGGTAC 4083  
Db 41211 ATGCGGAGCTGACGAGCGCTCCGGCGCGCTCGCGGCTGCGGCTGATCCGGCGCGGGTTCG 41270  
Qy 4084 CAGAGCCCGGTCCGCGCGAGTCTGAGCTGGAAGTGGCGCGCGGCGGTGGGTGAGCGGTTG 4143  
Db 41271 CGCGGACACCGGCTCGCTGCTGGAGCGCTCCGCGGAGCTGCGGTGGC-GATC 41329  
Qy 4144 CGCAGGGCGGAAGCGGGGAGGTGCGGCGGCGGAGCATATGGGGCGGCGGTCGAAA 4203  
Db 41330 CTCGCGTCTCAAGCGCGGCGCGGTATCTGCGGATCGATTCGCGGCGAGCGCGCGCG 41389  
Qy 4204 CGGCGCGCGGCGGTCCAGGCGTCTCCAGGAGCTAGACCGCGCAACCGCGCTTCG 4263  
Db 41390 CGCATCGCGGACATCGTGGCGGACCGCGCGCGCTGCTGCTGCGGCGGTTCAC 41449  
Qy 4264 CGCCAGGCGGCGGCGCGCGCGGCGGCTCGCGGTGTCGCGGTGCGCGCGCGCGGAG 4318  
Db 41450 GCCGAGCTGTTGGCGGCGCGCTCTCCGGGCTGCTGTCGCGCGCGCGGCTCGAG 41504

RESULT 10

US-11-052-554A-539/c  
; Sequence 539, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 539  
; LENGTH: 2196  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-539

Query Match 1.6%; Score 70.8; DB 8; Length 2196;  
Best Local Similarity 43.2%; Pred. No. 0.0078;  
Matches 677; Conservative 0; Mismatches 862; Indels 29; Gaps 6;  
Qy 2755 CAGCGACTACAGGACCTCGAACAACCGCTACCGGTGCTGCGAGACCATCTGCCCGC 2814  
Db 1990 CAACGCGCGCTTGGCGCGGTTGCCACCAACCGCTGCGGCGCGCGATACCGCTTCGCGCGT 1931  
Qy 2815 CATCGGGTTCCTCGGTGCTGCGCGCTCGATCTCTCCGAGGAGCGGAGGACGCGCT 2874

Db 1930 TTCGCGGTTCGCGCGGACACCGCGCGCGCGCATTTGCCGAAGAGCGACCCCGCGCGCGC 1871  
Qy 2875 CGCGCGGAACTGCGACCGACCGGGAAGGAC--TGACCGCGGAAGCGCCCGACACGGATA 2932  
Db 1870 GCGGAGGCGCGCGTCCGCGGCTTACGCGCAGTCCGCGCTGCTGCAACCGACAAAG 1811  
Qy 2933 CCGACCGCGGTTCGCGGCGCGACCGGGGCGCGCTACGAGGCGCGGCGCGGCGCGT 2992  
Db 1810 CGACACCGCGGTTCACCGCGCGGTCCGCGCGCGCGCGCTTGC CGCGCATGCGCGCGTGC 1751  
Qy 2993 AGTTCGCGCTGCGTACCGAAGCGTGGGGTCAAGAAATGTTCCGCGCTGTGCGATCGA 3052  
Db 1750 CGGTAAACCGCGCGTCCGCGCGCGCGCGCTTGC CGCGCATTCCGCGCGTTCGACCG 1691  
Qy 3053 CGTACTGCGCGGTATCCACCGTTCGAGTTCGAGAGCGCAGAAAGCGCACACAGTCGCGGA 3112  
Db 1690 TTGGGGGTTCGCGGTGCTGCGGTGCTGCTCCGCGCGCGCGCGCTGGCGCGC 1631  
Qy 3113 TGTCTCGGGTTCGCGATCGGTGGAACAGCGAGTTGGGGCGCAGTGC CGCGCGCGCT 3172  
Db 1630 CGGCGCGCGGTTCGCGT----TAGCAAGCGCGCATTTCCCGCGGTTCGCGCGCTCCCG 1575  
Qy 3173 CGGGGTTCGCGCGCGCGGTGCTTCATGTCGTTCCAGAAACCGCGCGCGCACCGGT 3232  
Db 1574 CCGTTCGCGCGATCGCGTCCGCGCGCTTTCGCGCGGTTCGCGCGCGCGCGCGCTCC 1515  
Qy 3233 TGACCGTATCCCGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCG 3292  
Db 1514 CCGCGGTGCGCGCGGTTCGATCAGCGGGCGTTCGCGCGCGCGCGCGCGCGCGCG 1455  
Qy 3293 CACCTTTGCTCATCGCGTATCCGATGGAATCGGGGAAGCGCGCGGTTCGCGCGGAGCG 3352  
Db 1454 GCGCGCGGT----CGTGCATTTGCGCGCGGATGTTGGCGCGCGCGCGCGCGCGT 1400  
Qy 3353 AGATGTTGATGATCGCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGCGCGCGCG 3412  
Db 1399 TCGCGCGTCCCGCGGTTCGCGTACAGTACACCGCGCGCGCGCGCGCGCGCGTTC 1340  
Qy 3413 CGGTTCGCGCGGTTCGCGCGCGCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTC 3472  
Db 1339 CGCGCGCGCGCGGTTCGCGCGCGCGCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCG 1280  
Qy 3473 TCCGTTCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCG 3532  
Db 1279 CGTTCGCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCG 1220  
Qy 3533 CGAATCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCG 3592  
Db 1219 CACCGTTCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCG 1160  
Qy 3593 CGGTTCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCG 3652  
Db 1159 TCGGTTCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCG 1100  
Qy 3653 CTGCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCG 3712  
Db 1099 CACCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCG 1040  
Qy 3713 GCGCGAGGCGGATACCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCG 3767  
Db 1039 AGAGCCACCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCG 980  
Qy 3768 CTTGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTC 3826  
Db 979 CGCGACCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTC 920  
Qy 3827 -----TGCGTGAGCGACCGCATGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTC 3877  
Db 919 CCAGGTACCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCG 860  
Qy 3878 CTTACTGTGACCGGTTCGAGCGGGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCG 3937  
Db 859 TCTCGGTGTTTTCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCG 800





Qy	219	CGCGGACGGATCACCCCTTGCGCTTCGGCCGAAGGCTTTCGCGTG--GTCCGCGCGCCAGAT	276
Db	29867	CGCGACGACGAAACGTTGCGACGACACCGGCGGCTGCGCGGGCGAGCGCGCGCAGAC	29926
Qy	277	CGCGAAGCGG-CTGGGGGGCGGCGCGTCACTTCGCGACACGCTCGGACGACACTTCGCGCT	335
Db	29927	CGCGTTCGCGCGAGCGCGCGGAGCCGATGTACCGGGTCGCGTGGAAGGACGTCGGCGCT	29986
Qy	336	TGCGCCCGCGCGCTGCCGCTGGCGCTCTCCAGGAAACGCGTCGCGGACGCGCTTCGGAT	395
Db	29987	GGAGGGGGCGACGTGGCGCGCCGAGAGACGCTGCTGCTCGCGGTGACGGTGGCTGCG	30046
Qy	396	ACTTTCGGAGCACTCTCGCGCGCGCTCCAGGCCCAAGTCTCTCGAAACGCAAGTGACC	455
Db	30047	GGCGCGCTGGGCGTTCGCGCGCGCGCGCGGCTGCGGAGCTGCTCGAGGCGCTG-GCGG	30105
Qy	456	GCCCGACGACCTCGGAGAGCGCGCGCTGTGCTTGGCGGTGATCGCTCGGGCCCGG	515
Db	30106	ACGCGCGCGCGCGCGCGCGCGCTGGTCTGTGGAACCTGACGCGCGCGCATGACGACGCGG	30165
Qy	516	ACAGCGGTACGCGCCGCTCCCTCGTGGCGCGGCGGCTCAGTCCCTGACCGCACTTCG	575
Db	30166	TCGTCGCGCGCTGACGCGCGGTCGCGCGCTGCGCCCTGCTGACGGGTGCTCG	30225
Qy	576	CGATGTGCGCGGATCGACGAGGCAACCGGGGACGTGCCGTACAGCGCGGACACGCG	635
Db	30226	CCGCGCGCAGCTGGCGGGGACGCGAGCTCTCTGTGTGACGCGCTTCGCGGTTGGGACCG	30285
Qy	636	CGTTCGACCGGATGGCGGGCGCCACGACAGCGTGTGGACATGAAGTCTCTGGCCCGCA	695
Db	30286	GGCGGACGAGGCGGTGACGCGCTGGGGCGCGCGCTCTGGGGCGCTGTGCGGCGCTGTCGGGGCCA	30345
Qy	696	GGAAGTCCAGTCTAGCCCGGACTCGCGTACGCGCGCTCGCTTCTCGCGCTGCGCGCGG	755
Db	30346	CGCGCGCGAG--TACCCGACCGCGGTCGCGGTCGCGGTCGACGTGGGCGCGGACCGCT	30403
Qy	756	TGATGAAGTCGTC-CGGCGCGGTTCCCCACGCGCGGATGGAAGCTTACACAGGTGC	814
Db	30404	GGAAGGGGCGCTCTTGCGTGGGCGCTGGCGCGGGGACGAGCGGAGCTTTCGCTGCG	30463
Qy	815	CGGACGCGGCTTCGCGCGCGCGCGGAAACGCTCGTCTCGG----CTCGGTGCG	870
Db	30464	CAGCGGAGGCGCGCGCGCTGCGGAGTGTGCGGAGCGAGCGCGCGCGGCG	30523
Qy	871	ACTGTCGTGACGAGGAACCGCGCGCGACCCCGTTGAGGGCCCGGTCCAGGCGCGGCGG	930
Db	30524	GCGGCGACGCGGTGATATCCGACGGAACAGCGCTGA--TACGCGGCGGACCGGGGAG	30581
Qy	931	GTGCGGTACTCGCCCGCGCAGACCTCGAGTTCGCGGCGGCTGACGGTCAACCGTTCGCG	990
Db	30582	CTCGGCGGCGATGTGCGGAAGCACTGTGTGAAGGCGCACGCGTGGCGGACCTCTGTGTG	30641
Qy	991	CCGCGCGCGGAGGACTCTGACGGGACCGGTCGCGGCGCAGAGTGGCGGACCTTGACGCGC	1050
Db	30642	ACGTCGCGCGCGGATGGAACGCGCCGACGCGCGCGCTGCTGTGACGAGCTGCGCGCG	30701
Qy	1051	G 1051	
Db	30702	G 30702	



Db 3005 GGGGGCCCTCGTCAGTTCAGACACAGCCGCTCGGGGCGCGCCCTTCTCCAGC 2946  
Qy 4219 TCCAGGGCTCCTCCAGACGCTTAGACCCGCAACCGCCCTGCGCAGCGGCGGACG 4278  
Db 2945 GCCTGAGCAGCTCCGGCAGCCCGCCAGCGCGAGCGGCCCCAGCGCCGAGCCAGCGCG 2886  
Qy 4279 GCCCGCCAGCTCCGCGCTGTCGGGTCCGGGGCCGAGG 4318  
Db 2885 CCGTCCGCGCCAGCACAGCTGGCGCTCGGCGGCCAGG 2846

RESULT 15

US-11-052-554A-532/c  
; Sequence 532, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 532  
; LENGTH: 2250  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-532

Query Match 1.5%; Score 66; DB 8; Length 2250;  
Best Local Similarity 43.7%; Pred. No. 0.044;  
Matches 642; Conservative 0; Mismatches 805; Indels 23; Gaps 7;  
Qy 2885 CTGGCACCGACCGGGAAGACTGACCGCGAAGCGCCCGCACCGGATACCGACCGCGCT 2944  
Db 2141 CGTTTACCGAGCCCGAAGCGGCTCCACCGGCGCGCGCTCCCGCGCGCTTCCCGCG 2082  
Qy 2945 GCCGAGCGCGGACCGGGGCGCTACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3004  
Db 2081 GCCCGCGCGCGCGCGCGCTTACCGCGCTTGGCGATCAGCCCGCGCTTCCCGCGCGCTCCG 2022  
Qy 3005 CGTACCGAAGCGTGG- CGGGTCAGAGATCGTTCCGCGCTGTGGCATCGACGTACTGGCGG 3063  
Db 2021 CAGCGCGCGCGAGTTCGGTTCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1962  
Qy 3064 GTGATCCACCGTGTGCTCGGAGCGCAGAAAGGCCACACGTCGGCGATGTCGTCGGGT 3123  
Db 1961 AACAGCAGCGCGCGCTCG 1902  
Qy 3124 CTGCGCATGCGGTTGAAACAGGAGTTGGGGCGGCGAGTCCGCG--GCCGCTCGGGGCT 3181  
Db 1901 CCGGCACCTCCGACCCCGCGGAGCGGACGAGCGCGCGGCTCGCGCGCGCGCGCGCGCG 1842  
Qy 3182 GCCCGCGCGCGTTCATGTCCTGCTCCAGAAACCGCGCGCGCGCGCGCGCGCGCGCGCG 3241  
Db 1841 GCCCGCGCGGTTCAAACCCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1782  
Qy 3242 TCCCGCGTTCCCGCAGTGTGCTGCGCGAGCGGCGAGCGTGTGCGTGTCCACCGCACCTTGG 3301  
Db 1781 GCGAGCAAGCCCGCTGCG 1724  
Qy 3302 TCATCGGTATCCGATGAGCTCGGGGAAAGCGCGCGCGCGGTGCGGCGAGACGAGATGTTGA 3361  
Db 1723 -----CGGATCCG 1670  
Qy 3362 TGATCCGCGCGCGGTGCG- GCAGTGTTCAGTCTGCTGAGACCGAGAAAGCGCGTGC- 3419

Db 1669 CGGCCCCCGGACGCTGTGTGCTCCGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1610  
Qy 3420 -CCGAGGTTGACGGCGGACGAGTCCGTCGAGAGACTCTCTCGGTGACTTCTCGTGTATCCGTC 3478  
Db 1609 GGCACCGCGCGCGCGCGCGCGCGCGCTCCCGCGCAGCGCGCGCGCGCTGACGCGCGACCGCG 1550  
Qy 3479 CCGAGCGCTGACGCGCGCGCGCTTTCACACAGAGATGTGAACTCTCGGGGGGCACTCCGAAC 3538  
Db 1549 CGGCGCGCGCGCGCGCGCGCGCTTGAACAGCCACCGCGCGCGCTTCGGGCGCGCGCGCG 1490  
Qy 3539 CGCCCATCCGCGCGTCGAACCGCGCTAGAGCGCGCGCGGTCAACCAAGACGCGCGGT 3598  
Db 1489 CGCGCGCGCGCGCGCGCGCGCTGAGACCGCGCGCGCTCCACCGCGCGCGCGCGCGCG 1430  
Qy 3599 CGGCGCGGATGGCCAAAGCCTGTCTCGCGCGCTGCTCCGATGGTCTCGACGCTCTCTCGG 3658  
Db 1429 CAAACAGCAAGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCCGCTGA 1370  
Qy 3659 CCGCGCGCTCGCTCTGCTGCTGCTGAGTGTGACTGTCAGAGCGCGCGCGCTCCGCGCGCGCG 3718  
Db 1369 CGGCG 1310  
Qy 3719 GGGCGATACCGCGCTCCGATGCCCGCGCTTCCCGCGTCCACAGGGCG-----GTCTGCG 3772  
Db 1309 CTCGCGCGCGCGCGCGCTGACGCGCGGACCGCGCGCGCTCCCTCCGCGCGCGCGCGCG 1250  
Qy 3773 CTTCCAGGGGTCTTCATACCTCTGCTCCATGTGACGATATCAGCCCCCGCGCGGTGCGT 3832  
Db 1249 CGCGCTGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTCGATCAGCG 1190  
Qy 3833 AGCGACCCATGCGCGCGCGCTCGCGCGCTTCGAATCGACGCTCAGACGCTTACCTGTGACCGCG 3892  
Db 1189 GCGCGCGCTCAGCGCTTGAATGCGCGCGTGTGATGCGGCGCAACGCTGTCTGCTGCGCG 1130  
Qy 3893 GTCAGACGGCGCGCGAGTGGCGCGTTCGACGCGTGGGGCGCAGATC---GGCGCGCGCG 3949  
Db 1129 TGTGACGCGGTTGCTGCG 1070  
Qy 3950 ACGGGGAAACCGCGCGCGTCAAGGGTCAAGGGTTCGCGGAGCGCGCGCGCGCGCGCGCG 4009  
Db 1069 CGATGCG 1010  
Qy 4010 CACCGACCGGATCGAGTTCGGCGCGTGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4069  
Db 1009 CGCGTTCGCGATCAGCG 950  
Qy 4070 ACTGTAAGCGTACAGAGCGCGTTCGGCGCGAGTCTGAGCTGAGCTGAGCTGCGCGCGCG 4129  
Db 949 CCGCGGTGGCAACCGACGTCGCGCGCTGCACCGTTCGCGCGCGTTCGCGATCAGCGCGCG 890  
Qy 4130 GGGTGAAGCGGTTGCGCGCGGCGGAGGCGGCGGAGTTCGCGCGCGGAGCATCATGGGCG 4189  
Db 889 CCG 830  
Qy 4190 GGGCGCGTCCGAAACCGCGCGCGCGCGCGCGTCCAGGGCTCTCCAGGACGCTAGACCG 4249  
Db 829 CCGACCG 770  
Qy 4250 CCAACCGCGCGTTCG 4309  
Db 769 CCGCGCGGAGTTACCGGTGTACCGAAGTTCGCGGAGCGCGCGCGCGCGCGCGCGCG 710  
Qy 4310 CGGCGCGGCGGAGCGCGGTAGAGCGCG 4339  
Db 709 CACCGGAGCGCGAAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 680

Search completed: January 19, 2006, 04:10:04  
Job time : 628 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2006, 16:18:41 ; Search time 3182 Seconds  
(without alignments)  
11294.380 Million cell updates/sec

Title: US-10-017-471B-19  
Perfect score: 4346  
Sequence: 1 gtcagcagcggtcggtggtt.....gctagagcgcggtacc 4346

Scoring table: IDENTITY\_NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	4346	100.0	4346	US-10-017-471A-19
2	401	9.2	401	Sequence 19, Appl
3	401	9.2	401	Sequence 12, Appl
4	377.2	8.7	9025608	US-10-017-471A-13
5	371.2	8.5	840	US-10-156-761-1
6	316.2	7.3	19016	US-10-469-442-1
7	274.2	6.3	88400	US-10-844-716-1
8	255.4	5.3	7600	US-10-732-859-1
9	253.6	5.8	9025608	US-10-156-761-1
10	250	5.8	2370	US-10-156-761-1965
11	236.8	5.4	906	US-10-156-761-924
12	229	5.3	699	US-10-049-710A-1
13	213	4.9	741	US-10-369-493-43039
14	210.6	4.8	762	US-10-369-493-28527
15	210.6	4.8	762	US-10-369-493-31286
16	206.6	4.8	750	US-10-369-493-35359
17	206.6	4.8	750	US-10-369-493-37943
18	206.6	4.8	750	US-10-369-493-38141
19	206.6	4.8	750	US-10-369-493-38559
20	185.4	4.3	732	US-10-369-493-35573
21	176.6	4.1	777	US-10-369-493-35593
22	174.6	4.0	762	US-10-369-493-37608
23	172.8	4.0	765	US-10-369-493-39254

c	24	172.8	4.0	765	6	US-10-369-493-39627	Sequence 39627, A
c	25	172.6	4.0	747	6	US-10-369-493-35639	Sequence 35639, A
c	26	171.4	3.9	747	6	US-10-369-493-35942	Sequence 35942, A
c	27	170	3.9	702	6	US-10-156-761-3692	Sequence 3692, Ap
c	28	169.2	3.9	753	6	US-10-369-493-40200	Sequence 40200, A
c	29	167.6	3.9	732	6	US-10-369-493-39993	Sequence 39993, A
c	30	164	3.8	726	6	US-10-369-493-30920	Sequence 30920, A
c	31	162.8	3.7	720	6	US-10-369-493-28161	Sequence 28161, A
c	32	162	3.7	765	6	US-10-369-493-24326	Sequence 24326, A
c	33	160.6	3.7	1029	7	US-10-437-963-98530	Sequence 98530, A
c	34	158	3.6	742	6	US-10-369-493-44176	Sequence 44176, A
c	35	154	3.5	699	6	US-10-369-493-32280	Sequence 32280, A
c	36	151.2	3.5	852	7	US-10-437-963-64696	Sequence 64696, A
c	37	151	3.5	663	6	US-10-156-761-2353	Sequence 2353, Ap
c	38	150.2	3.5	981	7	US-10-425-114-15642	Sequence 15642, A
c	39	150.2	3.5	1138	7	US-10-425-114-15867	Sequence 15867, A
c	40	150.2	3.5	1385	7	US-10-425-114-15463	Sequence 15463, A
c	41	150.2	3.5	1448	8	US-10-425-115-9925	Sequence 9925, Ap
c	42	148.4	3.4	819	6	US-10-156-761-4765	Sequence 4765, Ap
c	43	148.2	3.4	735	6	US-10-369-493-30930	Sequence 30930, A
c	44	148.2	3.4	780	6	US-10-369-493-28171	Sequence 28171, A
c	45	147	3.4	979	7	US-10-425-114-34802	Sequence 34802, A

ALIGNMENTS

RESULT 1  
US-10-017-471A-19  
; Sequence 19, Application US/10017471A  
; Publication No. US200301246441  
; GENERAL INFORMATION:  
; APPLICANT: Takano, Eriko  
; APPLICANT: Bibb, Mervyn  
; TITLE OF INVENTION: Antibiotic Production  
; FILE REFERENCE: 0380-P02329US1  
; CURRENT APPLICATION NUMBER: US/10/017, 471A  
; CURRENT FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: US 60/242,561  
; PRIOR FILING DATE: 2000-10-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 19  
; LENGTH: 4346  
; TYPE: DNA  
; ORGANISM: Streptomyces coelicolor  
US-10-017-471A-19

Query Match	100.0%	Score 4346;	DB 6;	Length 4346;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 4346;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GTGAGCAGCGGCGTGGGTTCCAGCCGACGGGTACTCGTTCCCGGCCACCGGGCACCG	60	
DB	1	GTGAGCAGCGGCGTGGGTTCCAGCCGACGGGTACTCGTTCCCGGCCACCGGGCACCG	60	
QY	61	GGTCTGCGCTCCATGACCGACCGCCATCGAGGACGTCGGCTGGCGGCTCTCTGATAGTGAC	120	
DB	61	GGTCTGCGCTCCATGACCGACCGCCATCGAGGACGTCGGCTGGCGGCTCTCTGATAGTGAC	120	
QY	121	GGCCCCCGCGCGGACGACATCGACGTCCTATCTCCCATCTGCGCCCCCGGAAAGTGAGC	180	
DB	121	GGCCCCCGCGCGGACGACATCGACGTCCTATCTCCCATCTGCGCCCCCGGAAAGTGAGC	180	
QY	181	ACCGCACCGCGGACGTCGACCCATGGGAGGCGCCATGCGGACCGGATCACCCTGGCT	240	
DB	181	ACCGCACCGCGGACGTCGACCCATGGGAGGCGCCATGCGGACCGGATCACCCTGGCT	240	
QY	241	TGCGCGGAAGGCTTCGCGTGGTCCGCCGCCACGATGCGGAACGGCTTGGCGGCGCGGCC	300	
DB	241	TGCGCGGAAGGCTTCGCGTGGTCCGCCGCCACGATGCGGAACGGCTTGGCGGCGCGGCC	300	
QY	301	CGTCACTTCCCGCAGCGTCGGCACGACCTGCGCCTTTGGCCCCCGCGCTGCGCTGGC	360	

Db 301 |||||CGTCACTTCCGCGACGCTCGGCAAGACTGCGCTTGCGCCCGCGCTCGCGTCCGC 360  
Qy 361 GCTCTCCAGGAACCGCTCGGCGAAGGCGCTCGGATATCTTCGGAAGCATCTGCTCGCGGC 420  
Db 361 GCTCTCCAGGAACCGCTCGGCGAAGGCGCTCGGATATCTTCGGAAGCATCTGCTCGCGGC 420  
Qy 421 CGCCTCCAGCCCGAGCTCTCGAAGCGAGTGAACCGCCCGAGCACCTCGGAGAGCGCGC 480  
Db 421 CGCCTCCAGCCCGAGCTCTCGAAGCGAGTGAACCGCCCGAGCACCTCGGAGAGCGCGC 480  
Qy 481 CGTCTGCTCGCTCGGCGGTGATGCTCGGCGCGGACAGCGCGTACGCGCGCTCCCTCGTG 540  
Db 481 CGTCTGCTCGCTCGGCGGTGATGCTCGGCGCGGACAGCGCGTACGCGCGCTCCCTCGTG 540  
Qy 541 GCCGGGCGCGGTCAAGTCCCTGAACGCCACTTTCGCGGATGCTCGCGCGATCGACGAGGC 600  
Db 541 GCCGGGCGCGGTCAAGTCCCTGAACGCCACTTTCGCGGATGCTCGCGCGATCGACGAGGC 600  
Qy 601 AACCGGGACGTGCGGTACAGCGGCGGACACAGCGCGTTCGGACCGGATGGGCGGCGCCA 660  
Db 601 AACCGGGACGTGCGGTACAGCGGCGGACACAGCGCGTTCGGACCGGATGGGCGGCGCCA 660  
Qy 661 GGACAGCGTGTGGACATGAAGTCTCGGCGCGAGGAAGTTCAGTCTAGCCCGGACTC 720  
Db 661 GGACAGCGTGTGGACATGAAGTCTCGGCGCGAGGAAGTTCAGTCTAGCCCGGACTC 720  
Qy 721 CGGTACGCGCCGCTGTTCTCGCGTGC CGCGGTGATGAAGTCTCGCGCGCGCTTC 780  
Db 721 CGGTACGCGCCGCTGTTCTCGCGTGC CGCGGTGATGAAGTCTCGCGCGCGCTTC 780  
Qy 781 CCCACCGCGAGCATGGAACAGTTTACAGGTGCGGAGCGCGGCTCGCGCGCGCGC 840  
Db 781 CCCACCGCGAGCATGGAACAGTTTACAGGTGCGGAGCGCGGCTCGCGCGCGCGC 840  
Qy 841 CGCGAAGCGCTGCTCGGCTCGGTGCGACTGTTCTGTCAGAGGAACGCGCGCGAC 900  
Db 841 CGCGAAGCGCTGCTCGGCTCGGTGCGACTGTTCTGTCAGAGGAACGCGCGCGAC 900  
Qy 901 CCCGTTGAGGCGCGGTCAGGCGCGGCTCGGCGTACTCGCGCGCGAGACTCGAC 960  
Db 901 CCCGTTGAGGCGCGGTCAGGCGCGGCTCGGCGTACTCGCGCGCGAGACTCGAC 960  
Qy 961 GTTTCGGGCGGTGACGCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1020  
Db 961 GTTTCGGGCGGTGACGCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1020  
Qy 1021 CCGGSCCAGAGTGGGCGACTGACGCGCGACACAGCGGTACGCGCGGTCAAGAAAT 1080  
Db 1021 CCGGSCCAGAGTGGGCGACTGACGCGCGACACAGCGGTACGCGCGGTCAAGAAAT 1080  
Qy 1081 CACTCGGGGCTCCTCTCGGCGAGCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
Db 1081 CACTCGGGGCTCCTCTCGGCGAGCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
Qy 1141 AGGGCAGATGTCGCGGCGGCGGAAACGCGGATGTTGCGGCGGCGGCGGCGGCTTC 1200  
Db 1141 AGGGCAGATGTCGCGGCGGCGGAAACGCGGATGTTGCGGCGGCGGCGGCGGCTTC 1200  
Qy 1201 AGCGGAGAACGCGGCGGCGGACGCTGAGGAAGAGAGGCTGCGGCGGCGGCGGCTTC 1260  
Db 1201 AGCGGAGAACGCGGCGGCGGACGCTGAGGAAGAGAGGCTGCGGCGGCGGCGGCTTC 1260  
Qy 1261 CGGTGACCGGACGCTGCTGACGCGCGCGCGCGCGGCGGCGGCGGCGGCGGATCC 1320  
Db 1261 CGGTGACCGGACGCTGCTGACGCGCGCGCGCGCGGCGGCGGCGGCGGCGGATCC 1320  
Qy 1321 AGCAGCGGCTGCGAATCTCCGCGTACCGGACGAAACCGGCGGCGGCGGCGGCGG 1380  
Db 1321 AGCAGCGGCTGCGAATCTCCGCGTACCGGACGAAACCGGCGGCGGCGGCGGCGG 1380  
Qy 1381 AGGGCGCGGACCGGTACAGGCGGCGCGGCTGCGGCTGCGGCTGCGAGCAGCATGC 1440

Db 1381 AGGGCGCGGACCGGTACAGGAGCAGCGCGCTGCGCGCTCGAGCAGCAGCATGC 1440  
Qy 1441 CCGGTACGTGCTGTTGGGCGCTTGGGAAGGCTCGGCTGACTGCTGCACCCGCGAGTC 1500  
Db 1441 CCGGTACGTGCTGTTGGGCGCTTGGGAAGGCTCGGCTGACTGCTGCACCCGCGAGTC 1500  
Qy 1501 GCCACGTGCTCTGCTGCGAACTCGCCGACAGGACACGTCCTCGCTCGCGGTGCGACCG 1560  
Db 1501 GCCACGTGCTCTGCTGCGAACTCGCCGACAGGACACGTCCTCGCTCGCGGTGCGACCG 1560  
Qy 1561 CGCGCGCGCGGCGACGCGCGCGCTCCGGGCAACGATGCGGTGGAGTTCGCGAGTCGC 1620  
Db 1561 CGCGCGCGCGGCGACGCGCGCGCTCCGGGCAACGATGCGGTGGAGTTCGCGAGTCGC 1620  
Qy 1621 CGCGCATCCCGCGGTAGACTTGAAGACTGTTGAAGCGGCTGTTGAAGCGGCTCCCGTGGCAG 1680  
Db 1621 CGCGCATCCCGCGGTAGACTTGAAGACTGTTGAAGCGGCTGTTGAAGCGGCTCCCGTGGCAG 1680  
Qy 1681 CGAGCCGTTCGCGCGCGGCGACGCGCCCACTGCTGCTGCTGACCGGCTGCCCGCGC 1740  
Db 1681 CGAGCCGTTCGCGCGCGGCGACGCGCCCACTGCTGCTGCTGACCGGCTGCCCGCGC 1740  
Qy 1741 GGAATTTCACTGCGGAAACAGGCCACTTCCACTCCAGCTCCGCGACCTCGCCGACAGC 1800  
Db 1741 GGAATTTCACTGCGGAAACAGGCCACTTCCACTCCAGCTCCGCGACCTCGCCGACAGC 1800  
Qy 1801 CGAGTGTGTCGAGTGGCAGGTGATGTCAGCGTGGCCATCAGGAAGTGGTAGCCACCG 1860  
Db 1801 CGAGTGTGTCGAGTGGCAGGTGATGTCAGCGTGGCCATCAGGAAGTGGTAGCCACCG 1860  
Qy 1861 GCACGCGTAGCCGCGTGGGAAGCAGGACATCGCGCTCGACGAGGCTCTCGCGCATCA 1920  
Db 1861 GCACGCGTAGCCGCGTGGGAAGCAGGACATCGCGCTCGACGAGGCTCTCGCGCATCA 1920  
Qy 1921 GCACGCGTAGCTGTCGCTGCGGTCGCGTGGAAAGACGCGTGGTCTGTTGGGCGAGGA 1980  
Db 1921 GCACGCGTAGCTGTCGCTGCGGTCGCGTGGAAAGACGCGTGGTCTGTTGGGCGAGGA 1980  
Qy 1981 CGGCGGTGACCGAGAACCGGTCACCGGCTTCGCTGATCCAGCTGACCGGGAACCGTCT 2040  
Db 1981 CGGCGGTGACCGAGAACCGGTCACCGGCTTCGCTGATCCAGCTGACCGGGAACCGTCT 2040  
Qy 2041 GCACCTGCTGCTCGGTGACAGCGCCATCGGAACCGGCAATCGCGTTCGATCGAGT 2100  
Db 2041 GCACCTGCTGCTCGGTGACAGCGCCATCGGAACCGGCAATCGCGTTCGATCGAGT 2100  
Qy 2101 TGGCATCGGACCGAGAAATGATCAAACTA CTGCTTCGCGGATGCGGTCCTCCCGCAGGAAT 2160  
Db 2101 TGGCATCGGACCGAGAAATGATCAAACTA CTGCTTCGCGGATGCGGTCCTCCCGCAGGAAT 2160  
Qy 2161 CATGTGATCGGAGCTGTTCTGTATGCGCGAACGTTAAGATACAGACTGAGCGGTTTTT 2220  
Db 2161 CATGTGATCGGAGCTGTTCTGTATGCGCGAACGTTAAGATACAGACTGAGCGGTTTTT 2220  
Qy 2221 TTCTATCTTCTCCCGGGGAGACATGAACAAGAGCGGCGATGCGCAAGAGGACCGGCG 2280  
Db 2221 TTCTATCTTCTCCCGGGGAGACATGAACAAGAGCGGCGATGCGCAAGAGGACCGGCG 2280  
Qy 2281 GATCCGCAACGCGGAGACGATCTTCGAGCGCGCGCGAGGTCTTCGAGAAAGCGGCTA 2340  
Db 2281 GATCCGCAACGCGGAGACGATCTTCGAGCGCGCGCGAGGTCTTCGAGAAAGCGGCTA 2340  
Qy 2341 CCAAGCTGCGACGATCAGGAGATCCTCAAGGTGGCGGGGTGACCAAGGAGGCGCTCTA 2400  
Db 2341 CCAAGCTGCGACGATCAGGAGATCCTCAAGGTGGCGGGGTGACCAAGGAGGCGCTCTA 2400  
Qy 2401 CTTTCACTTCCAGTCCAAGGAAGAACTGGCGCTGGCGCTCTTCGACGCGCCAGGACCC 2460  
Db 2401 CTTTCACTTCCAGTCCAAGGAAGAACTGGCGCTGGCGCTCTTCGACGCGCCAGGACCC 2460  
Qy 2461 ACAGCGCTTCGAGGACAAACCCCTCGGCTGCGAGAACTCATCGACATGGGCGATGTT 2520  
Db 2461 ACAGCGCTTCGAGGACAAACCCCTCGGCTGCGAGAACTCATCGACATGGGCGATGTT 2520



2521 CTGTCAACCGCTTGCAGCAAGACGTCGTGTGCCCCGGGCGGCGTGTGCGCTTCTCATGGAACA 2580  
2521 CTGTCAACCGCTTGCAGCAAGACGTCGTGTGCCCCGGGCGGCGTGTGCGCTTCTCATGGAACA 2580  
2581 GCAGGGCAGCGCTCTGATCGCGGAGGACCTTCCGTGCGTGGCAGGACACTCTCTGAA 2640  
2581 GCAGGGCAGCGCTCTGATCGCGGAGGACCTTCCGTGCGTGGCAGGACACTCTCTGAA 2640  
2641 GCTGCTGAACAGGCGCCAGGAGAACGCTGAGTTGCTGCCCCATGTGTCACACCGACTC 2700  
2641 GCTGCTGAACAGGCGCCAGGAGAACGCTGAGTTGCTGCCCCATGTGTCACACCGACTC 2700  
2701 GCGCGATCTTACGTGGGCAAGCTTCCGCGGGATACAGGTGCTGTGCCAGACGGTCAGGAA 2760  
2701 GCGCGATCTTACGTGGGCAAGCTTCCGCGGGATACAGGTGCTGTGCCAGACGGTCAGGAA 2760  
2761 CTACCAAGGACCTGGAACACCGCTACCGCTGCTGCGAAGACATCTCTGCCCGCCATCGC 2820  
2761 CTACCAAGGACCTGGAACACCGCTACCGCTGCTGCGAAGACATCTCTGCCCGCCATCGC 2820  
2821 GGTTCCTCCGTGCTGGCGCGCTCGATCTCTCCGAGAGCGGAGCAGCGCTCGCGGC 2880  
2821 GGTTCCTCCGTGCTGGCGCGCTCGATCTCTCCGAGAGCGGAGCAGCGCTCGCGGC 2880  
2881 CGAACTGGCAACCGGAGGAGACTGACCGCGGAGCGCCGCAACCGGATACCGACCGG 2940  
2881 CGAACTGGCAACCGGAGGAGACTGACCGCGGAGCGCCGCAACCGGATACCGACCGG 2940  
2941 CCGTCCCGAGCGCGGCGGCGGCGCTACCGGCGCGGCGGCGGCGGCGGCGGCTAGCTTCG 3000  
2941 CCGTCCCGAGCGCGGCGGCGGCGCTACCGGCGCGGCGGCGGCGGCGGCGGCGGCTAGCTTCG 3000  
3001 CTGTGCTACGGAAGCGTGGCGGCTCAGAGATGCTTCGCGCTGTGTCATCGACGACTCGG 3060  
3001 CTGTGCTACGGAAGCGTGGCGGCTCAGAGATGCTTCGCGCTGTGTCATCGACGACTCGG 3060  
3061 CCGGTGATCCACCGTGGAGTGGTGGAGGCGCAGAAAGGCCACCAAGTGGCGGATGCTGC 3120  
3061 CCGGTGATCCACCGTGGAGTGGTGGAGGCGCAGAAAGGCCACCAAGTGGCGGATGCTGC 3120  
3121 GGTCTGCGATGCGGTTGAACACGAGGTTGGGCGGCGAGTGGCGGCGGCGGCGGCGGCTC 3180  
3121 GGTCTGCGATGCGGTTGAACACGAGGTTGGGCGGCGAGTGGCGGCGGCGGCGGCGGCTC 3180  
3181 TCGCGCGCGGCTGCTGATGCTCCAGAAACCGGCGGCGGCGGCGGCGGCGGCGGCTG 3240  
3181 TCGCGCGCGGCTGCTGATGCTCCAGAAACCGGCGGCGGCGGCGGCGGCGGCGGCTG 3240  
3241 ATCCCCCGTTCCCGCCAGTTGCTGGCCAGGCGGAGCGTGAAGGTTGTCACCGACCGCTTG 3300  
3241 ATCCCCCGTTCCCGCCAGTTGCTGGCCAGGCGGAGCGTGAAGGTTGTCACCGACCGCTTG 3300  
3301 GTCATCGGATATCCGATGATGCTCGGGGAAACGCGCGCGGCGGCGGCGGCGGCGGATGTTG 3360  
3301 GTCATCGGATATCCGATGATGCTCGGGGAAACGCGCGCGGCGGCGGCGGCGGCGGATGTTG 3360  
3361 ATGATCCGCGCGGCTGCGGAGTGGTTCAGTCCGTCGTCGACGAGGAAACGCGGTCG 3420  
3361 ATGATCCGCGCGGCTGCGGAGTGGTTCAGTCCGTCGTCGACGAGGAAACGCGGTCG 3420  
3421 CGGACGCTTGAACGCGGACGAGTGGTTCGAGACCTCTCTCGGTGACTTCGCGTATCCGTCGC 3480  
3421 CGGACGCTTGAACGCGGACGAGTGGTTCGAGACCTCTCTCGGTGACTTCGCGTATCCGTCGC 3480  
3481 GAGCGCGTGAACGCGGCTGTTTACAGAGTGTGAACTCGGCGGCGGCGGCGGAACTCG 3540  
3481 GAGCGCGTGAACGCGGCTGTTTACAGAGTGTGAACTCGGCGGCGGCGGCGGAACTCG 3540  
3541 CCCATCCCGGCTGGAACCGCGGCTAGAGCGGCGGCGGCTGCAACGAGCGGAGTTGCG 3600  
3541 CCCATCCCGGCTGGAACCGCGGCTAGAGCGGCGGCGGCGGCTGCAACGAGCGGAGTTGCG 3600

3601 GCCCGGATGGCCAAACGCTGTGTCGCGGCTGTCTCCGGATGGTCTCGACGGTCTCTCGCGCC 3660  
3601 GCCCGGATGGCCAAACGCTGTGTCGCGGCTGTCTCCGGATGGTCTCGACGGTCTCTCGCGCC 3660  
3661 GCCGCTCGCTGTCTGCTAGTGGAGTGTCCACGAGCGGCGGCGGCGGCGGCGGCGGCGG 3720  
3661 GCCGCTCGCTGTCTGCTAGTGGAGTGTCCACGAGCGGCGGCGGCGGCGGCGGCGGCGG 3720  
3721 GCGATACCGGCTCGATGCGCCCGGCTTCCCGCGGTCAACAGGGCGGTCTTGGCCCTCCAGC 3780  
3721 GCGATACCGGCTCGATGCGCCCGGCTTCCCGCGGTCAACAGGGCGGTCTTGGCCCTCCAGC 3780  
3781 GGTCTTCCATACCTCGTCCCATGTGTCAGCATATCAGCCCCCGGCGGCTGCGTGAAGCGACCC 3840  
3781 GGTCTTCCATACCTCGTCCCATGTGTCAGCATATCAGCCCCCGGCGGCTGCGTGAAGCGACCC 3840  
3841 ATGGCGCGGCTCGGCGCGTTTCCGAATCGACGGTCAACAGCTTCTGTGACCGGCTCAGAGCG 3900  
3841 ATGGCGCGGCTCGGCGCGTTTCCGAATCGACGGTCAACAGCTTCTGTGACCGGCTCAGAGCG 3900  
3901 GGGCCGAGTGGCCCGGTTGGACGCGTGGGGCCAGATCGGGCGGCGCGCACGGGGAAACCG 3960  
3901 GGGCCGAGTGGCCCGGTTGGACGCGTGGGGCCAGATCGGGCGGCGCGCACGGGGAAACCG 3960  
3961 GGGCCGAGTGGGGGTGAGGGTTCGCGGGACCGCCAGCGGCGGTCAAGGCAACCGACCGGA 4020  
3961 GGGCCGAGTGGGGGTGAGGGTTCGCGGGACCGCCAGCGGCGGTCAAGGCAACCGACCGGA 4020  
4021 TCGAGGTGGGGGTGCGCACGCGGCGCACAGTCTCTCGCGGCGGAGTCCGAGCTCGTACGCG 4080  
4021 TCGAGGTGGGGGTGCGCACGCGGCGCACAGTCTCTCGCGGCGGAGTCCGAGCTCGTACGCG 4080  
4081 TACCAAGACCGGTCGCGCGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 4140  
4081 TACCAAGACCGGTCGCGCGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 4140  
4141 TTGCGCCAGGGCGGAGGCGGGGAGGTGCGGCGGAGCATCATGCGGCGGCGGCGGCGGTCG 4200  
4141 TTGCGCCAGGGCGGAGGCGGGGAGGTGCGGCGGAGCATCATGCGGCGGCGGCGGCGGTCG 4200  
4201 AAACGCGCGGCGGCGGCTCTCCAGAGCTCTCCAGAGCTCTAGACCCCGCAGCGGCGCC 4260  
4201 AAACGCGCGGCGGCGGCTCTCCAGAGCTCTCCAGAGCTCTAGACCCCGCAGCGGCGCC 4260  
4261 TCGCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4320  
4261 TCGCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4320  
4321 AGCGAGCGGTAGAGCGGCGGCGGTACC 4346  
4321 AGCGAGCGGTAGAGCGGCGGCGGTACC 4346

RESULT 2

US-10-017-471A-12  
; Sequence 12, Application US/10017471A  
; Publication No. US2003012464A1  
; GENERAL INFORMATION:  
; APPLICANT: Takano, Eriko  
; APPLICANT: Bibb, Mervyn  
; TITLE OF INVENTION: Antibiotic Production  
; FILE REFERENCE: 0380-P02329US1  
; CURRENT APPLICATION NUMBER: US/10/017,471A  
; CURRENT FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: US 60/242,561  
; PRIOR FILING DATE: 2000-10-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Streptomyces coelicolor  
US-10-017-471A-12

Query Match 9.2%; Score 401; DB 6; Length 401;  
Best Local Similarity 100.0%; Pred. No. 8.7e-86;  
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1973 GGGCAGGACGGCGGTGACCGAGAACCGGTACCGGCCCTTCGGTATCCAGCTGACCGGAA 2032  
DB 1 GGGCAGGACGGCGGTGACCGAGAACCGGTACCGGCCCTTCGGTATCCAGCTGACCGGAA 60

QY 2033 CGCGTCTCGACCCCTGGTCCGGTGGACAAAGCGCCATCGGAACCGGCAATGCGGTTGTTC 2092  
DB 61 CGCGTCTCGACCCCTGGTCCGGTGGACAAAGCGCCATCGGAACCGGCAATGCGGTTGTTC 120

QY 2093 GATCGAGTTGGCATCGGACGAGCAATGATCAAACTACTGCTTCGGGCATGGTCCCCC 2152  
DB 121 GATCGAGTTGGCATCGGACGAGCAATGATCAAACTACTGCTTCGGGCATGGTCCCCC 180

QY 2153 CCAGGAATCATGTATGCGGAGCTGTTCTGTATGCGGAAGCTTAAAGATACAGACTGAGC 2212  
DB 181 CCAGGAATCATGTATGCGGAGCTGTTCTGTATGCGGAAGCTTAAAGATACAGACTGAGC 240

QY 2213 GGTGTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGGATGCCCAAGCAG 2272  
DB 241 GGTGTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGGATGCCCAAGCAG 300

QY 2273 GACCGGGCGATCCGACGCGGAGACGATCTTGAAGCGCGGCGGAGGCTTCGAGAG 2332  
DB 301 GACCGGGCGATCCGACGCGGAGACGATCTTGAAGCGCGGCGGAGGCTTCGAGAG 360

QY 2333 CAGGGCTACCAAGCTGCCACCATACGAGAGTCTCAAGGT 2373  
DB 361 CAGGGCTACCAAGCTGCCACCATACGAGAGTCTCAAGGT 401

## RESULT 3

US-10-017-471A-13/c  
; Sequence 13, Application US/10017471A  
; Publication No. US20030124644A1  
; GENERAL INFORMATION:  
; APPLICANT: Bibb, Mervyn  
; APPLICANT: Takano, Eriko  
; TITLE OF INVENTION: Antibiotic Production  
; FILE REFERENCE: 0380-P02329US1  
; CURRENT APPLICATION NUMBER: US/10/017,471A  
; PRIOR FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: US 60/242,561  
; PRIOR FILING DATE: 2000-10-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Streptomyces coelicolor  
US-10-017-471A-13

Query Match 9.2%; Score 401; DB 6; Length 401;  
Best Local Similarity 100.0%; Pred. No. 8.7e-86;  
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1973 GGGCAGGACGGCGGTGACCGAGAACCGGTACCGGCCCTTCGGTATCCAGCTGACCGGAA 2032  
DB 401 GGGCAGGACGGCGGTGACCGAGAACCGGTACCGGCCCTTCGGTATCCAGCTGACCGGAA 342

QY 2033 CGCGTCTCGACCCCTGGTCCGGTGGACAAAGCGCCATCGGAACCGGCAATGCGGTTGTTC 2092  
DB 341 CGCGTCTCGACCCCTGGTCCGGTGGACAAAGCGCCATCGGAACCGGCAATGCGGTTGTTC 282

QY 2093 GATCGAGTTGGCATCGGACGAGCAATGATCAAACTACTGCTTCGGGCATGGTCCCCC 2152  
DB 281 GATCGAGTTGGCATCGGACGAGCAATGATCAAACTACTGCTTCGGGCATGGTCCCCC 222

QY 2153 CCAGGAATCATGTATGCGGAGCTGTTCTGTATGCGGAAGCTTAAAGATACAGACTGAGC 2212

DB 221 CCAGGAATCATGTATGCGGAGCTGTTCTGTATGCGGAAGCTTAAAGATACAGACTGAGC 162

QY 2213 GGTGTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGGATGGCCCAAGCAG 2272

DB 161 GGTGTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGGATGGCCCAAGCAG 102

QY 2273 GACCGGGCGATCCGACGCGGAGACGATCTTGAAGCGCGGCGGAGGCTTCGAGAG 2332

DB 101 GACCGGGCGATCCGACGCGGAGACGATCTTGAAGCGCGGCGGAGGCTTCGAGAG 42

QY 2333 CAGGGCTACCAAGCTGCGACGATCACGAGAGTCTCAAGGT 2373

DB 41 CAGGGCTACCAAGCTGCGACGATCACGAGAGTCTCAAGGT 1

## RESULT 4

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 8.7%; Score 377.2; DB 6; Length 9025608;  
Best Local Similarity 64.1%; Pred. No. 2.8e-80;  
Matches 568; Conservative 0; Mismatches 318; Indels 0; Gaps 0;

QY 227 GATCACCCCTGGCTTCGGCCGAAAGGCTTCGGGTGGTCCGCGGCCCGCCAGATCGGGAACGCGC 286

DB 4362444 GAGCATGCCGCATTTCAGGCGAAGCGGTTCTGTGTGTCGTCGCCCTGTGCTAGGAG 4362385

QY 287 CTGGCGGGCGCGCCGCTCACCTTCGCGACGCTTCGCGACGACTGCGCTTCGGCCCCCGCC 346

DB 4362384 GGTGTGGTGGCCCGTGACCCGGCGGAGACGTGGGGTGGTGTGCTGTTCGGGCCCAGC 4362325

QY 347 CGCTGCCCTCGCGGCTCTCCAGGAACGCGTTCGGCGACGCGGCCCTCGGATACTTCGAGAC 406

DB 4362324 AGCTGGCGTTGTGCGCTCAGGACGAGTGCCTGCACCACTTTGGGCGGTAGCGCGCC 4362265

QY 407 ATCTGCTCGCGCGCGCTTCAGCCCGCAGCTCTCTGGAACGAGTGAACGCGGCCCGCCAGCAC 466

DB 4362264 CACCGGGCGGAGCGCTGCTGTGGTGTTCAGTTTCGCGGAGACGAGCGGGCGCGGAGGACC 4362205

QY 467 TCGGAGACCGCGCGCTGCTGCTGCGCGTGTATCGCTTCGGGCCCGGACAGCCCGCTGAC 526

DB 4362204 GCGCCAGGTCCGGTCTGTTCCTCCGAGAGACCGCTTCGGGGCCGCTGAGGGGATAC 4362145

QY 527 GCCCGTCCCTGTGCGCGCGCGGTTCAGTGCCTTCAGCCCGCACTTCGCGGATGTTCGCGC 586

DB 4362144 ACCCGGCCCTCGTGCGCGGGTGAAGTCAAGCGCGGTCAACGCGACCTCGGCGATGTACGG 4362085





```
; TYPE: DNA
; ORGANISM: Streptomyces sp. H021
US-10-469-442-1

Query Match      7.3%; Score 316.2; DB 9; Length 19016;
Best Local Similarity 61.3%; Pred. No. 1.3e-65;
Matches 529; Conservative 0; Mismatches 328; Indels 6; Gaps 1;

Qy 2922 GCGACCGGATACCGACCGCGCGTGCCTCCGAGCGCGCGCGCGCTACCGCGCGCGC 2981
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8079 CGACTCGGTTCGGATGCCGGGTTCGGAACCGCGCGGTCCGACACACGTCGGCTCGGA 8138
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2982 GCGCGCGCGTAGTCTGCCCTGCGTACCGAAGCGTGCAGAGATCGTTCGCGC 3041
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8139 CCGCTCGGGTTCGACCGCTCGGGTTCGGAACCGGTGCGGAGAGTGTCCGCC 8198
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3042 TGTGGCATTCGACGTAATCGCGCGGTGATCCACCGTGAAGTCTCGGAGGCCAGAAAGGCCAC 3101
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8199 GGTGGCGTCGATGAAGGAGCGGTGATCCAGGGCGGTCTCGGAGGCGAGGACGCCAC 8258
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3102 CACGTCCGCGATGTCGCGGTCTGCCGATCGGTGTAACAAGGATTTGGCGCGCAGTGC 3161
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8259 CACGTACCGACGTCGCGCGGTTCGCCACCTGTCCGAAGCGCGACATCTGGCGCATCTG 8318
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3162 CGCGCGCGCTCGGGGTCTGCCCGCGCGTGGTTCATGTCCGTCTCCAGAAACCGG 3221
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8319 CTCACCGCGCTCGCGGATGTGAAACACCGGGTGCCTGTGCGGT-----GATCCCGGG 8372
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3222 CGCACCGCGTTGACCGGTGATCCCGGTTCGCCAGTTGCTGCGGAGGCGAGCGTGAG 3281
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8373 GCGGACGTGTTCAGGTGATGTCGCCGGCGCCAGGTATTGCGGAGTGCAGGCGAT 8432
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3282 CGTGTCCACCGACCTTTGGTCAATCGCTATCCGATGCACTCGGGGAAACGCGCGCGGT 3341
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8433 CTGTTCCGACCGCGCTTTGGTCAATCGCTAGCGCGCTCTTTCGGGGTTGGCGAAACCGGGT 8492
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3342 CGCGCAGACGATGTTGATGATCCGCGCGCGTCCGCGAGTCTGTTCACTTCGTGCTG 3401
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8493 GAGCCCGAGGAGATGTTGATGATGTCGGCGCGCGTTCGCGAGCAGTCCGAGCGCGCTG 8552
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3402 GACCAGGAACGCGTGCCTCCGAGCGTTCACGCGCGACAGTCCGTCCGAAGACTCTCTCGGT 3461
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8553 GACGATGAAGAACGCGCTTTCGCTTGAATGGCGAAGAGCGGTTCGACAGTCCGGTGT 8612
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3462 GACTTCGTGATCCGTCGCGAGCGCTGACGCGCGCGTGTTCACGAGATGTGAATC 3521
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8613 GACTCTCTCCGGGGGACGCGCGCCATGATGCGCGCTTGTGACAGGATGTGAGGGT 8672
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3522 GGGCGGCACTCCGAATCCGCGTCCGCGTTCGAACCGCGGTAGAGCGCGCGCGCTC 3581
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8673 GCGGACCCCGGTTCGTTCTTCACTCCGCTCTCCAGCGCTTCGAACAGCGTGTCAATGTC 8732
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3582 ACCCAGCAGCGCGAGTTCGGCGCGGATGGCCAAACGCTGTTCGCGCGCTGCTCCGATGTT 3641
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8733 GCGGGGAGCGCGAACTCGCTCGAAGCGCGAAACGCTCGGCGCGCCCGCGCGCTCAATGTC 8792
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3642 CTCGACGCTCTCTCGCGCGCGCTTCGCTGCTGCGGTAGTGAATGCAAGAGCGCGCC 3701
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8793 GCGGACCGTGTCTACGCGCGCGTCTTCGCGGTGCGGTAGTGCAGCGGCACCACTGCGCC 8852
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3702 GTCCCGCGGACCGCAGGCGATACCGGTTCGATGCGCGCGCTTCCCGCGTCAACAG 3761
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8853 CTCATGGGCGAGCCGAGTGGCGGTTCGCGCGTCCGATGCCCGGTTCGACCCCGGTACAG 3784
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3762 GSCGCTTTGCTCCAGCGGTTC 3844
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 8913 TCGGCTCTTGTTCGAGGCTTTC 8935
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 7

US-10-844-716-1  
; Sequence 1, Application US/10844716  
; Publication No. US20050003409A1

```
; GENERAL INFORMATION:
; APPLICANT: Huang, Chengjin
; APPLICANT: Chaleff, Deborah T.
; APPLICANT: Ruppen, Mark E.
; APPLICANT: Stephens, Jerome
; TITLE OF INVENTION: Cloning Genes From Streptomyces Cyaneogriseus Subsp.
; TITLE OF INVENTION: Nonycanogenus for Biosynthesis of Antibiotics and Methods of Use
; FILE REFERENCE: AM100484
; CURRENT APPLICATION NUMBER: US/10/844,716
; CURRENT FILING DATE: 2004-05-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 88400
; TYPE: DNA
; ORGANISM: bacteria
US-10-844-716-1

Query Match      6.3%; Score 274.2; DB 8; Length 88400;
Best Local Similarity 54.2%; Pred. No. 1.2e-55;
Matches 689; Conservative 0; Mismatches 543; Indels 39; Gaps 5;

Qy 2541 ACGTCTGTGGCCCGGCGCGGTGCGCTCTCCATGAGACACGAGCGACGCTCTCATC 2600
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11188 AGGACTCGCCCGCGCGCGGTGCTCCAGGACGAGGAACCTGCTGATCGCGCGTGC 11247
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2601 GCCGAGAACCTTCCGTGCTGCGTGGCAGAGACACTCTGA-----AGCTGCTGAAC 2654
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11248 GAACCGACCTGCGGGTCTCCAGGACGAGGAACCTGCTGATCGCGCGTGC 11307
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2655 CCAAGGAGAACGGTGAAGTGTGCTGCCCATGTGGTCACACCGACTCGGCGCGATCT 2714
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11308 GCGCGGACTACCGGCGCATCGGACCTACCGCGCGCGGAGACGCGGTGCTGCGT 11367
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2715 TGGGACAGTTGCGCGGATACAGTGTGTCAGAGAGGTGAGGACTACAGAGACCTG 2774
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11368 CGGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11427
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2775 AAGACCGGTACGCGTCTGCGAGAGACACATCTCTCCGCGCGCGCTCGGTTCC 2834
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11428 GGAGCGGCCACACACCGCGGAGTCCGAGGTGCTACCTTCTCGGCGGGGACTTCT 11487
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2835 TGGCGCGCTCGATCTCTCCAGGAGCGGAGACGCTCTCGGCGCGAACTGGGACCG 2894
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11488 TCGAGCGCCACACGACGCGGTGTG-GAGTCTGCTACCGCGCGCTCGGAGACCG 11546
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2895 CCGGGAAGGACTGACCGCGGAAGCGCGGACACCGGATACCGACCGCGCTCGGAGCG 2954
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11547 GCGCCCGCGCGACCGGGTGTGAGGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 11606
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2955 CCGACCGGGCGCGCTACCGGCGCGCGCGCGCGCGCTAGTCTGCTGCTGCTACCG 3014
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11607 TCTGCGGACCG--GGCGGACCGGACCCCGCGCGGACAGCGCGCGGACCTCAT 11664
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3015 CGTGGCGGTTCAGAGAACTGTTCCGCTGTGGCATCGACGTACTGGCGCGGTGATCC 3074
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11665 CATGGCGGCGCGGCGCATTCGCCCGCTGCGGTCCAGCAGTTGGCGGCGGTGATCC 11724
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3075 TGAGTCTGCGAGGCGCAAGAAAGGCGACCACTGCGGCGATGCTGCGGCTGTCGCG 3134
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11725 GCGTCTGCGAGACACGAGAGGCGACGATGCGGCGGATGTCTGTCGCGCGCGCGCG 11784
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3135 GTTGAACAGAGTGTGGCGCGCATGTCGCGCGCGCGCTCGGCGGTCTGCGCGCGCG 3194
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11785 GCCGAGCGGTTCAGGGCGGAGATGCCCTTGGCGCGCGCGCGCGCGCGCGCGCG 11844
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3195 GTTCATGTCGTCTCCAGAAACCGCGCGCGCACCGCTTGGACCGGTGATCCCGGTCCC 3254
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11845 GTTCATGTCGTCTCCAGAAACCGCGCGCGCACCGCTTGGACCGGTGATCCCGGTCCC 11904
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3255 CAGTTGCTGCGCGGCGAGCGGTGAGGTGTCCACCGCACCTTGGTTCATCGCGTATCC 3314
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11905 GAGTTCTGCGAGCGCGGAGCGCATCATCTCCAGCGCGCGCGCTTGGTTCATGGCGTA 11961
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```











```
Db 524 ACGAGGTAATGACCGCGCCCTTGGTGGCGCATAGGCGGAGCGGTGCGGACCGGGG 465
Qy 3337 CCGGTGCGGCGACAGAGATGTTGATGATCCCGCGCGCTCGCGAGTGGTTTCAGTCGG 3396
Db 464 CCGCGCGGAGCGAGAGAGACACCGCTGCTGCTTGCATACATACCGGAGCAAC 405
Qy 3397 TGCTGGACAGAGAACAGCGGTGCGCGGAGAGTGGAGCGGAGACAGTGGTGGAGACCTCC 3456
Db 404 TGCTGACAGAGAGAGTGGGTGCGGCGACGTTTACGCGGCAACAGATCGTGGAGTCTTCC 345
Qy 3457 TCGGTGACTTCTCGATCGTCCGAGCGCTGAGCGCGCGCTGTTTACACAGGATGCG 3516
Db 344 ACGGTGCTGCTCGATGCTGGCGCTTCGAAATCCCGGCGTTCGCCACCAAGATATCC 285
Qy 3517 AACTCGGCGGCACTCCGAACTCCGCAATCCCGGCGTTCGAAAGCGCGGTAGAGCGCGCC 3576
Db 284 AGCGGTG-----GCCGATCACCGCGGCACTTCGCGGCGGCTGCTGTCG 240
Qy 3577 GGTGACCCACAGCGCGAGTTCGGCGCGGATGGCCAGACCTGTCGCGCGCTGCTCGG 3636
Db 239 GGGCGGTGCGGCTCGCGAGATTCGCGGAGATTTCTGCGCATTCGCGCGCGCGCGGA 180
Qy 3637 ATGGTCTGACAGGTCTCTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3696
Db 179 ATTCGCGGACCAACGCGCATCCGCTTCTTCTGCGCATTCGCTGAGTGAACCAAGACCTGT 120
Qy 3697 GCGCGTCCGCGCGGCGAGCGGCGATACCGCGTCCGATGCGCGGCTTCCCGCGTCC 3756
Db 119 GCGCGGCTTCGCGAGCGGAGTGGCTGGCGGCGCGATGCGGCGTGAAGGCGCGTG 60
Qy 3757 ACCAGGCGGTCTTGCC 3773
Db 59 ACGAGCGGTCTTGCC 43
```

## RESULT 15

```
US-10-369-493-31286/c
; Sequence 31286, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31286:
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
US-10-369-493-31286
```

```
Query Match 4.8%; Score 210.6; DB 6; Length 762;
Best Local Similarity 57.7%; Pred. No. 2.2e-40;
Matches 425; Conservative 0; Mismatches 294; Indels 18; Gaps 2;

Qy 3037 CCGCTGTGGCATCGAGTACTGGCGGTGATCCAGCGTGGTGGTGGGCGGCGGAGAG 3096
Db 761 CCACCGTGCATGCAAGTCTCGCGGTAACTCAGCGTGGCGGTCCGAGGCGAGAG 702
Qy 3097 GCCACCACTGCGCGATGCTGCGGTCTGCCGATGCGTGGTGAACACGAGTGGCGGCG 3156
Db 701 GCACGCGCGCGCATATCATCCGCTGGCGCACCGGCTTGAGCGCTGCGATGCGGAGC 642
Qy 3157 AGTGGCGCGCGCGCTCGGGGTCTCGCGCGGTGCGGTTCATGCTCCAGGAA 3216
```

Search completed: January 19, 2006, 03:59:39  
Job time : 3191 secs

this page blank (uspto)

this page blank (uspto)

this page blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2006, 10:27:16 ; Search time 709 Seconds  
(without alignments)  
10896.027 Million cell updates/sec

Title: US-10-017-471B-19  
Perfect score: 4346  
Sequence: 1 gtcagcagcgctcggtt.....gcgtagcgcgcggtacc 4346

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:  
1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*  
8: /cgn2\_6/ptodata/1/ina/RB COMB.seq:\*  
9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	255.4	5.9	7584	3	US-09-446-681-1 Sequence 1, Appli
C 2	255.4	5.9	7584	3	US-09-446-681-2 Sequence 2, Appli
C 3	255.4	5.9	7600	3	US-09-469-211A-1 Sequence 1, Appli
C 4	219.8	5.1	14638	3	US-09-902-540-1106 Sequence 1106, Ap
C 5	216	5.0	744	3	US-09-902-540-2423 Sequence 2423, Ap
C 6	184.6	4.2	912	3	US-09-489-039A-6755 Sequence 6755, Ap
C 7	161.4	3.7	906	3	US-09-489-039A-3950 Sequence 3950, Ap
C 8	159	3.7	792	3	US-09-252-991A-11222 Sequence 11222, A
C 9	159	3.7	936	3	US-09-252-991A-11165 Sequence 11165, A
C 10	156.8	3.6	876	3	US-09-252-991A-12001 Sequence 12001, A
C 11	155.4	3.6	879	3	US-09-252-991A-11911 Sequence 11911, A
C 12	150	3.5	798	3	US-09-489-039A-430 Sequence 430, App
C 13	145.6	3.4	30001	2	US-08-125-468-1 Sequence 1, Appli
C 14	145.6	3.4	30001	2	US-08-474-933-1 Sequence 1, Appli
C 15	139.6	3.2	1571	3	US-09-902-540-336 Sequence 336, App
C 16	139.2	3.2	741	3	US-09-902-540-6232 Sequence 6232, Ap
C 17	136	3.1	780	3	US-09-252-991A-12374 Sequence 12374, A
C 18	136	3.1	954	3	US-09-252-991A-12328 Sequence 12328, A
C 19	126	2.9	18469	3	US-09-902-540-1205 Sequence 1205, Ap
C 20	123.6	2.8	579	3	US-09-252-991A-11983 Sequence 11983, A
C 21	122.4	2.8	5467	3	US-09-902-540-703 Sequence 703, App
C 22	119.6	2.8	750	3	US-09-902-540-7356 Sequence 7356, Ap
C 23	119.2	2.7	666	3	US-09-252-991A-12234 Sequence 12234, A
C 24	118.8	2.7	804	3	US-09-252-991A-15809 Sequence 15809, A

ALIGNMENTS

RESULT 1

US-09-446-681-1/c  
; Sequence 1, Application US/09446681  
; Patent No. 6849442  
; GENERAL INFORMATION:  
; APPLICANT: Archer, John AC  
; APPLICANT: Summers, David K  
; APPLICANT: Roland, Herve J  
; APPLICANT: Powell, Justin AC  
; TITLE OF INVENTION: Biosensor materials and methods  
; FILE REFERENCE: 0380-P02083-US0  
; CURRENT APPLICATION NUMBER: US/09/446,681  
; PRIOR FILING DATE: 2000-03-14  
; PRIOR APPLICATION NUMBER: PCT/GB98/01893  
; PRIOR FILING DATE: 1998-06-29  
; PRIOR APPLICATION NUMBER: GB 9713666.7  
; PRIOR FILING DATE: 1997-06-27  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 7584  
; TYPE: DNA  
; ORGANISM: Rhodococcus corallina  
US-09-446-681-1

Query Match 5.9%; Score 255.4; DB 3; Length 7584;  
Best Local Similarity 60.6%; Pred. No. 5.3e-35;  
Matches 454; Conservative 0; Mismatches 291; Indels 4; Gaps 2;

QY	3020	CGGGTCAGAGATCGTTCCGCTCTGGCATCCAGTACTGCGCGGTGATCCACCGGAGT	3079
DB	7584	CTGGTTAGACCGGTTGGCGTTCGATGACCTGCCCGGTGATCGACCGCGG	7525
QY	3080	CGTCGGAGCCAGAAAGCCACACGTCGGCATGTCGGGTCTGCCGA---TGCGGT	3136
DB	7524	CGCGCTGACGAGGAGGCCAGATCGCGCGCATGTCCTCCGGGTGGCGAGTTTCGCA	7465
QY	3137	TGAACACGAGTGGCGGCAGTGGCGCGCCCTCGGGGFTCTCCCGCGCGGTGCGT	3196
DB	7464	GTGCGTGTGGACGCGCGGTGGTGGCGGCATGTCGTCACCGAAGCAGTGGCGGT	7405
QY	3197	TCATGTCGCTCCACGAACCGCGCGCCACCGCGTGTACCGGTGATCCCGGTTCGCCA	3256
DB	7404	TCATGTCGCTATCGAGCGCGCGCGCCACCGCGTGTGGCGGTGATGCTCTGTTCCCGA	7345
QY	3257	GTTCCTGCGCCAGGCGGAGCGGTGTCCACCGCACCCCTTGGTTCATCGCGCTATCCGA	3316

Db 7344 CGTCTACGCGGAGGCGCGGGTGAGACCTCGATCGCCCCCTTGCTCATGGCGTAGTCGA 7285  
Qy 3317 TGGACTCGGGGAAACCGCGCGCGGGTCGCGGCAGACGAGATGTTGATGATCGCCCGCGGT 3376  
Db 7284 TGAAGTCGGGTCTGGCGTAGCGGGCGGATCCGGAGGAAATGTTGAACGATGCGACCGCGGT 7225  
Qy 3377 CGCGCAGTCGTTTTCAGTCCGTGCTGAGCAGAGAAACAGCGGGTGCCGCGAGCTTGACGGCGA 3436  
Db 7224 CGTGATCCGGGGCAGGGCATGCCGAGTCACGAGAAACGGGGCGGCTGGTTGATGTCGA 7165  
Qy 3437 CCAGTCGCTCGAAGACCTCTCTCGTGAATTCCTGATCCGTCGATCCGAGCGCTGACGCGCG 3496  
Db 7164 CCAGACGGTTCGAAATCTCTGACAGTGAACGCTCGAGCGCTCGCGCGCGACTGATTCGCG 7105  
Qy 3497 GCTTGTTCACAGATGTCGAATCGGGGGGCACTCCGAACTCGCCATCCCGGGTTCGA 3556  
Db 7104 GCTTGTTCACAGGATGTCGAAGCCCTCGGTTCGAGCCCGAGACCGCTCGAGCGCGGAGTTCGA 7045  
Qy 3557 ACGCCGCTAGAGCGCGCGCGCTCACCCACGAGCCGAGTTCGGCCCGGATGGCCAAACG 3616  
Db 7044 ACTCCGCGATCAGCTCTTACGCCCCCTCGGGTCGGGACAGGTTCGGCTTGGAACCGCAGCG 6985  
Qy 3617 CCTGTCGCGCGTGTCTCGGATGCTCTGACAGGTCTCTCGCGCGCGCCCTCGCTGCTGC 3676  
Db 6984 CGAGCCCCCGCAGCGCTGATGCGCTCGAC-ACGCGACGCGCGCGCGCTCGATCGGATC 6926  
Qy 3677 GGTAGTGAATGCGACAGCGCCCGCTCGCGCGCGCGCGCGAGGGCGATACCGGCTTCGA 3736  
Db 6925 GGTAGTGAATGCGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6866  
Qy 3737 TGCCCCCGCTTCCCGCGTTCACAGGCG 3765  
Db 6865 TGCCGCGCTGCGCCCGGTGACACGCG 6837

## RESULT 2

US-09-446-681-2  
; Sequence 2, Application US/09446681  
; Patent No. 6849442  
; GENERAL INFORMATION:  
; APPLICANT: Archer, John AC  
; APPLICANT: Summers, David K  
; APPLICANT: Roland, Herve J  
; APPLICANT: Powell, Justin AC  
; TITLE OF INVENTION: Biosensor materials and methods  
; FILE REFERENCE: 0380-P02083-US0  
; CURRENT APPLICATION NUMBER: US/09/446,681  
; CURRENT FILING DATE: 2000-03-14  
; PRIOR APPLICATION NUMBER: PCT/GB98/01893  
; PRIOR FILING DATE: 1998-06-29  
; PRIOR APPLICATION NUMBER: GB 9713666.7  
; PRIOR FILING DATE: 1997-06-27  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 7584  
; TYPE: DNA  
; ORGANISM: Rhodococcus corallina  
US-09-446-681-2

Query Match 5.9%; Score 255.4; DB 3; Length 7584;

Best Local Similarity 60.6%; Pred. No. 5.3e-35;  
Matches 454; Conservative 0; Mismatches 291; Indels 4; Gaps 2;  
Qy 3020 CGGGTCAGAGATCGTTCGCGCTGTGGCATCGACGTACTGGCGGTGATCCACCGTGAGT 3079  
Db 1 CTGGTTAGACCGGTTGCGTTGGTGGCGTGCATGACCTGCCCGGTGATCGACCGCGG 60  
Qy 3080 CTTGCGAGCGCAAGAGCCACACGTGCGCGATGTCGTGGGTCTGCCGA---TGCGGT 3136  
Db 61 CGGCGCTGACGAGGAGGCCACGATCGCGCGATGTCCTCGCGGTGGCGAGTTTTCGA 120  
Qy 3137 TGAACACGAGTTGGCGGCGAGTGC CGCGCGCTCGGGGTCTGCGCGCGCGCGCGTGC 3196

Db 121 GTGCAGTGGTGACGCGCGGTGTCGGGGCATGGTCTGTCACCGAAGCCAGTCGCGGT 180  
Qy 3197 TCATGTCCGTCTCCAGAAACCCGCGCGCACCGGTTGACCGTGTATCCCGGTTCCCGCA 3256  
Db 181 TCATGTCCGTATCGAGCGCGCGCGCCACGCGGTTTGGCGGTGATGCTCTGTCGCGGA 240  
Qy 3257 GTTCCCTGGCCAGGCGGAGGTGTCACCGCACCTTGGTTCATGCGGTATTCGA 3316  
Db 241 CGTCTACGCGAGGCGCGGTGAGCACCTCGATCGCCCCCTTGGTTCATGCGGTAGTCGA 300  
Qy 3317 TGGACTCGGGAAACCGCGCGCGGTTCGCGGACGAGATTTGATGATCCCGCGCGT 3376  
Db 301 TGACGTCCGGTCTCGGTAGCGGCGGATCCGAGGAAATGTTGACGATGCGACCGCGGT 360  
Qy 3377 CGCGAGTCTGTTTCAGTCCGTGCTGAGACAGGAAACAGGGTTCGCGGAGTTCGAGCGGA 3436  
Db 361 CGTGCATCCGCGGCGAGGCGCATGCCGAGTTCACGAAACCGGGCGCGCTGGTTCGAGTTCGA 420  
Qy 3437 CCAGTCGCTCGAAGACCTCTCGGTGACTTCGCTGATCCGTCGCGAGCGGTGACGCGCG 3496  
Db 421 CCAGACGCTCGAAATCTCTGACAGTCGCGCTCGAGCGCTCCGCGCGCATGATTCGG 480  
Qy 3497 GCTTGTTCACAGGATGTCGAACTCGGGCGGCACTTCGAACTCGCCCATCCCGCGGTTCGA 3556  
Db 481 GCTTGTTCAGCAGATGTCGAGCCCTCGGTGAGCCCGGAGACCGTCGAGCGCGGAGTTCGA 540  
Qy 3557 ACGCGCGTAGAGCGCGCGCGCTACCCACGACCGCGAGTTTCGCGCGCGATGCGCAACG 3616  
Db 541 ACTCCCGCATCAGCTCTTTCAGGCGCCCTCGGGTCGGGACAGGTTCGCGCTGGGCTGGAGCCG 600  
Qy 3617 CTTGTCGCGCGCTCTCGGATGTCGAGCGTCTCTCGCGCGCGCGCTCGCTGCTGC 3676  
Db 601 CGAGCCCCCGCGAGCGCTGATGCGCTCGAC-ACGCGAGCGCGCGCTCGATCGGATC 659  
Qy 3677 GGTAGTGAATGCGACGAGCGCGCGCTCGCGGCGAGCGCGCGCGCGCGATACCGGTTCGA 3736  
Db 660 GGTAGTGAATGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 719  
Qy 3737 TGCCCCCGCTTCCCGCGGTTCACAGGCGG 3765  
Db 720 TGCCGCGCTGCGCCCGGTGACACGCGG 748

## RESULT 3

US-09-469-211A-1/c  
; Sequence 1, Application US/09469211A  
; Patent No. 6660524  
; GENERAL INFORMATION:  
; APPLICANT: J. Turck  
; APPLICANT: J. Archer  
; TITLE OF INVENTION: CONTROL OF GENE EXPRESSION IN EUKARYOTES  
; FILE REFERENCE: 9341-021  
; CURRENT APPLICATION NUMBER: US/09/469,211A  
; CURRENT FILING DATE: 1999-12-22  
; PRIOR APPLICATION NUMBER: UK 9828660.2  
; PRIOR FILING DATE: 1998-12-24  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 7600  
; TYPE: DNA  
; ORGANISM: Rhodococcus sp.  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (295)..(1035)  
; OTHER INFORMATION: ohpR regulator  
; NAME/KEY: CDS  
; LOCATION: (1261)..(2805)  
; OTHER INFORMATION: ohpA transport  
; NAME/KEY: CDS  
; LOCATION: (2807)..(4720)  
; OTHER INFORMATION: ohpB monooxygenase

NAME/KEY: CDS  
LOCATION: (5721)..(6665)  
OTHER INFORMATION: ohpD catechol 2,3-dioxygenase  
US-09-469-211A-1

Query Match  
Best Local Similarity 5.9%; Score 255.4; DB 3; Length 7600;  
Matches 454; Conservative 0; Mismatches 291; Indels 4; Gaps 2;

3020 CGGGTCAGAGAAATCTTCGCTGTGGCATCGACACTACTGGCCGGTGATCCACCGTGAGT 3079  
|||||  
7595 CTGGTTAGAGCCGGTTGCGTTGGTGGCTCGATGACCTGCCCGGTGATCGACCGCGG 7526  
3080 CTTGCGAGCCAGAAAGCCACACGTCGCGCATGTCTGCGGTCTGCCGA---TGGCGT 3136  
|||||  
7525 CGCGCTCAGCAGGAAGGCGACGATCGCGGGCATGTCTCTCGCGGTGGCGAGTTTCGCA 7466  
3137 TGAACACGAGTTGCGCGCCAGTGC CGCGCGCTCGGGGTCTGCGCGCGCGCTGGT 3196  
|||||  
7465 GTGCAGTGGTGAACCGCGCGGTGGTGGCGGCATGTCTGTCACCGGAAGCCAGTCCGGT 7406  
3197 TCATGTCCGTCTCCACAGAAACCGCGCGCCACCGCGTTGACCGTGATCCCGCTTCCGCCA 3256  
|||||  
7405 TCATGTCCGTATCGAGCGCGCGCGCGCCACCGCGTTGGCGGTGATGCTGCTGTTCCCGA 7346  
3257 GTTGCCTGCGCAGGCGGAGCGTGAAGTGTCCACCGCACCTTGGTCTATCGCGTATCCGA 3316  
|||||  
7345 CGTCTACGCGAGGCGCGGGGTGAGCACCTCGATCGCCCCCTTGGTCTATGGCGTAGCTGA 7286  
3317 TGGACTCGGGGAAACCGCGCGCGGGTGGCGGACGAGATGTTGATGATCCGCCCGCGT 3376  
|||||  
7285 TGACGTGGGTCTGGCGTAGCGGGCGGATTCGGAGGAATGTTGACGATGCCACCGCGT 7226  
3377 CGCGCAGTCTGTTTCACTGCTGACACAGGAAACAGCGGTGTCGCGGAGTGTGACGGCGA 3436  
|||||  
7225 CGTGATCGGGGCGAGGATGTCGAGTACAGAGAACGGGGCGGCTGTTGAGTGGA 7166  
3437 CAGTGGTTCGAAGACCTCTCGGTGAATTCGCGTGAATTCGTCGCGAGCGGTGACGCCG 3496  
|||||  
7165 CCAGACGGTCGAAATCTCGACAGTGAACGCGCTCGAGCGCTCCGCGCGGACTGATTCGG 7106  
3497 CTTGTTTACACAGGATGTGCACTCGGGCGGCACTCCGAACTCGCCCATCCCGGCGTGA 3556  
|||||  
7105 CGTTGTGACGAGGATGTGACGCTTCGAGCTTCGAGCCCGAGACCGTCGAGCGGGAGTGA 7046  
3557 ACGCCGCTGAGAGCGCGCGCTCACCAACGACGCGCGAGTTCGCGCCCGATGGGCAACG 3616  
|||||  
7045 ACTCCGATCAGCTTTTCAGGCGCTTCGGGTGCGGACAGGTTCGGCTGGACCGCGCG 6986  
3617 CTTGTCCGCGCTGCTCCGATGGTTCGACGGTCTCTCGCGCGCGCGCTCGCTGCTGC 3676  
6985 CGAGCGCCCGGCGAGCGGTGATGCGTGCAC-ACCGCAGCGCGCGCGTCCGATCGATC 6927  
3677 CGTAGTGAATCCACGAGCGCGCTCGCGCGGACGCGAGGCGGATACCGCTCCGA 3736  
6926 CGTAGTGAACGATTACCGCGGACCGAGGCGGCGCACCGCATCTCGATGGCGGCCCGCA 6867  
3737 TCCCGCGCTTCCCGGCTCACAGGGCG 3765  
6866 TGCCGGGCTGCGCCCGGTGACAGCGCG 6838

RESULT 4  
US-09-540-1106/c  
; Sequence 1106, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B

CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 1106  
LENGTH: 14638  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-540-1106

Query Match  
Best Local Similarity 5.1%; Score 219.8; DB 3; Length 14638;  
Matches 429; Conservative 0; Mismatches 327; Indels 3; Gaps 1;

3009 CCGAAGCGTGGCGGTTCAGAGAAATCGTTCCGCTGTGGCATCGACGTTACTGCGCGGTGAT 3068  
12645 CTGGAGCGCGCGCGCTCAGAGCATCATCCCGCCGAGGCGCTCGATGCGTGGCCCGTCA 12586  
3069 CCACCGTGAAGTCTCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3128  
12585 CCACCGTCTCCGCGGACAGCAGCATGCGGACCATCCCGCATCCCGCATCCCGCATCC 12526  
3129 GATGCGGTGAACACGAGGTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3188  
12525 CACAGCGCGAGCGCAACTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 12466  
3189 CCGTGCCTTC--ATGTCGCTCTCCAGAAACCGCGGCGGCGGCGGCGGCGGCGGCGGCG 3245  
12465 GACGCGCGCGCGGAGTCCGTCGCGGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 12406  
3246 CCGTTCCTCCGAGTTGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3305  
12405 CCGGCGCGCGGAGTCTTTCGCGAGGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 12346  
3306 CCGGTATCCGATGAGTCCGCGGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3365  
12345 CCGGTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 12286  
3366 CCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3425  
12285 GCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 12226  
3426 GTTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3485  
12225 GTGAGCGTTCATCAGCGGCTCGAACTGGGCGCTCGGTCGTCCTCCGGAAGTGA 12166  
3486 GCTGACCGCGCGGTTGTTTACCGAGATGTCCGAACTCGGCGGCGGCGGCGGCGGCGG 3545  
12165 GCGCATGCCCGGTTTTCACCGAGAGTCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12106  
3546 CCGCGCGTGAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3605  
12105 CTTCCGTCCGCGCGCTTCCAGGAGCGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCG 12046  
3606 GATGCGCAACCGCTGTCGCGCGCTGCTCGGATGCTCTCGAGGTCCTCTCGCGCGCGG 3665  
12045 GAGCGCGCGCGCTTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11986  
3666 CTCGCTGCTGCGTAGTGAAGTCCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 3725  
11985 CCGCGCGGTGCGGTAGCTCAGGATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 11926  
3726 ACCCGGTCCGATGCGCGCGGCTTCCCGCGGTACACAGGCG 3764  
11925 GTTCCGCGCGGCTCCAGCGGCTTCCCGCGGTGATGAGAGC 11887

RESULT 5  
US-09-540-2423/c  
; Sequence 2423, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B

APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(115949)B  
CURRENT APPLICATION NUMBER: US/09/502,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 2423  
LENGTH: 744  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-502-540-2423

Query Match 5.0%; Score 216; DB 3; Length 744;  
Best Local Similarity 56.6%; Pred. No. 2.5e-28;  
Matches 421; Conservative 0; Mismatches 320; Indels 3; Gaps 1;  
QY 3024 TCAGAGATCGTTCCGCTGTGGCATCGACGTACTTGGCCGGTGTATCCACCGTGTGTC 3083  
DB 744 TCAGAGATCATCCGCGCCGAGGCTCGATCGCTGGCCCGTCAACCGAGCGCTCTCCG 685  
QY 3084 GGAGCCGAAAGGCCACCGTCTGGCGATGTCTCGGGTCTGCGGATGCGGTTGAACAC 3143  
DB 684 CGACAGCATGTGCGACCGCATCCCGATGCTGTGGGCTGACCCACAGCGCGAGCGC 625  
QY 3144 GGAGTGTGGCGCAGTGTGCGCGCGCTCGGGGTCTGCGCGCGCGTGTGTC--AT 3200  
DB 624 AACCTGGCGCGCAGCTTGTTCAGTCTCGGGTTGTACGGAGGACGCGCCCGCAA 565  
QY 3201 GTCCGTCTCCAGAAACCGCGCCACCGCTGTGACGTGATCCCGCTTCCCCAGTTG 3260  
DB 564 GTCCGTGCGGTGCGCGCCCGCGGAGCACGTTCAGCGAAATCTCCGGGCGCCAGCTC 505  
QY 3261 CTGGCGAGGCGAGCGTGTGAGCGTGTCCACGCACTTGTGTCATCGGATTCGATGGA 3320  
DB 504 CTTCGCGAGGTACCGGGTGAGCGCTCCACGCGCGCTTCATGTGCGGTAGGCGGCGT 445  
QY 3321 CTCGGGAAACCGCGCGCGGTGCGGAGACAGATGTGATGATCGCCCGCGTCCG 3380  
DB 444 GCGAGGAGGTGAGCGCGCGAGCGCGCGAGGATGTTCAGATGCGCCCGCATCCG 385  
QY 3381 CAGTGTTCAGTCTGAGTCCAGAACAGCGGTGCCCGAGTGTGACGGGACCG 3440  
DB 384 CAGCACCGGCAACAGCGCTGCGTGAGGAGAACTGTCCTTGGAGTGAAGTTCATCAG 325  
QY 3441 TCGGTGGAAGACTCTCTCGGTGACTTTCGTGATCCGTCCGAGCGGTGACCGCGGTT 3500  
DB 324 CGCGTCGAACCTGGGCTCGGTCTCGCGCAAGCTGACTGGATGCGCCATGCGCGGTT 265  
QY 3501 GTTACACGAGTGTGGAACCTGGGCGGCACTCCGAACCTCGCCATCCGCGGTGGAACG 3560  
DB 264 GTTACACGAGAGTCCAGCGCTCGCGCCGAGTGAAGTGAAGCGCCAGCTCTCGTCCG 205  
QY 3561 CGCGTAGAGCGCGCGGTGACCCACAGCGCGGATTCGCGCGGATGGCCAAAGCGCTG 3620  
DB 204 TTCCAGAGAGCGCGAGCCCGGTGTCTCCCGACGTGAGCGGAGCGCCACCGCTT 145  
QY 3621 TCGCGCGTGTCTCGGATGTGTGACGCTCTCTCGCGCGCGGTCTGCTGCTGCTGCTA 3680  
DB 144 GCGCGCTGTCTTCAATCTGCTTTCAGACCGCGCGCTTCTGTCGCGCGGTGCGGTA 85  
QY 3681 GTGAGTCCAGAGCGCGCTCGCGCGCGAGCGCGAGGCGGATACCGCGTCCGATGCC 3740  
DB 84 GGTAGGATGATGCGCGGTGCGCGCGCGCGCGCGCGCTTGTGAGCGGCTGTTCCGCG 25  
QY 3741 CCGGCTTCCCGGTCACAGGCG 3764  
DB 24 ACGGCTTCCCGCGTATGAGAGC 1

RESULT 6  
US-09-489-039A-6755/C  
Sequence 6755, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 6755  
LENGTH: 912  
TYPE: DNA  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-6755

Query Match 4.2%; Score 184.6; DB 3; Length 912;  
Best Local Similarity 53.5%; Pred. No. 6.2e-23;  
Matches 410; Conservative 0; Mismatches 354; Indels 3; Gaps 1;  
QY 3024 TCAGAGATCGTTCCGCTGTGGCATCGACGTACTTGGCCGGTGTATCCACCGTGTGTC 3083  
DB 912 TCAGATGTTGTTCGCGCCGAGACCTCGATCCGCTGGCGGTACCCAGCGGTTATCGTC 853  
QY 3084 GGAGCCGAGAAAGGCCACCGTCTGGCGATGTCTCGGGTCTGCGGATGCGGTTGAACAC 3143  
DB 852 GCGTAGCAGGCTGCGCATCATCGGCCCAATCTCGGAAACGCCACCGCTCCAGCGC 793  
QY 3144 GGAATTGCGGCGAGTGTGCGCGCGCTCGGGGTCTGCGG---CGCGGTGTTTCAT 3200  
DB 792 CGTCATGCGCGCAACTCGCGCTTTACTCTGCGTCTGCGCGCACCGACCGCCGCAAA 733  
QY 3201 GTCCGTCTCCACGAAACCGCGCGCACCGGTTCACCGTGTATCCCGTTCGCCCGAGTTG 3260  
DB 732 ATCGTGGCGATAGCGCGCGGGCGATGTATTCGATGATGCGCGCGCGCGCGCGTGC 673  
QY 3261 CCTGGCAGGCGAGCGGTGAGCGTGTCCACGCACTTGTGTCATCGGTATTCGATGGA 3320  
DB 672 GCGGCCATGTACAGCTCAGCATCTCCACAGCGCTTTGGCGCGCGGTAGGCGGAGAA 613  
QY 3321 CTCGGGAAACCGCGCGCGGTGCGGCGAGACGATGTTGATGATCCGCGCGCGTCCG 3380  
DB 612 GCCAGGATAGGATACGCGGCTCAGCCCGGAGGAAAGTTAACGATCCGCGCGCGTCTGC 553  
QY 3381 CAGTGTTCAGTCTCGTCTGAGCACAGGACAGCGGTCCCGGACGTTGACGCGGACCGAG 3440  
DB 552 CAGCAGCGGCGAGCGGTTTGTACAGGAAATATACCCCTTAACTGAGATGTTAAACAG 493  
QY 3441 TCGGTGGAAGACCTCTCTCGGTGACTTCGTGATCCGTCCGAGCGGTGACCGCGCGTT 3500  
DB 492 CCCGTGCAACTCGCTTCGCGCTTCGCGCGAGCGGGCGGAAATTCGCGTCCGCTGCGTT 433  
QY 3501 GTTACACGAGTGTGAACTCGGGCGGACCTCCGAACTCGGCCATCCCGGCGTCCGAGCG 3560  
DB 432 ATTGATCAGGTGTGAAAGTGGCGCGCGCCCGCCAGACGCTGGCCAGCGCGCTGCCAGG 373  
QY 3561 CGCGTAGAGCGCGCGCTCACCGACGCGCGAGTTTCGCGCGGATGGCGCAACCGCGT 3620  
DB 372 GTCGCGAGGCGCGGAGCTGCGGCTTGTGCGCATATCCAGCGGCGAGGCGATGCGCTT 313  
QY 3621 TCCGCGCTGTCTCGGATGTTCTCGACGTTCTCTGCGCGCGCGCTCTGCTGCTGCGCTA 3680  
DB 312 CGGCGCCAGGCGCGGATGTGCGGAGCACCGCTTCCGCTGTTTCGAGACTGCGCGGTA 253  
QY 3681 GTGAGTCCGACGAGCGCGCTCGGGCGAGCGCGGAGGCGGATACCGCGTCCGATGCC 3740  
DB 252 GGTGAGATGACATCCCGACCTCGCGGCGGATATCCAGTGTGCGGTCTGACGCGCAAGGCC 193







Db 426 TCCTGATACCCCGCCCGCTCGCCCATGTGTTGACCGCCCGCTTGGTGGGACGAGA 485  
Qy 3416 GTGCCCGGACGTTGACGGCGACAGTCGGTCAAGACCTCTCTCGGTGACTTCGGTGATCC 3475  
Db 486 CGCTGGCAGGTTGATCGCCAGGCTCGCTCGAAGTGGCCAGCGGAGCTCTTCCACCG 545  
Qy 3476 GTCCGAGCGCTGACCGCCCGGTTGTTTACAGAGNTGCAACTCGGGCGGCACTCCGA 3535  
Db 546 AACCGAGGCGAGGACCGCGGGGTTGTTGACAGGATGTGATGCG-----TCCGA 596  
Qy 3536 ACTCGCCCATCCCGGCTGCAACGCGCTAGAGCGCGCCGCTCACCCAGACGCGCA 3595  
Db 597 AGCGAGCCCGCTTCTGTCGAGCGGCGGCGAGCGCGCCCGCT-----CGGCG 647  
Qy 3596 GTTCGCGCCGATGCGCAACGCTGTCTCCCGCGTGTCTCGGATGTCTCGACGCTCTCTC 3655  
Db 648 TATCGCATGACAGGCGCAGGATGCTGGCTCCGGCGCTTCGAGTTCCGGCGCCAGGGCC 707  
Qy 3656 GCGCGCGCCCTCGCTGCTGCTGCTAGTGAATGCTGCAAGCGCGCCGCTCGGGCGGCGCC 3715  
Db 708 GCGACTGTCTCGCCGAGCTGAGTGTGAGGCGGAGCGCTGCGCGCTTCGGGCGGCGCC 767  
Qy 3716 GCAGGCGGATACCGGCTCGGATGCGCGGCTTCCCGGCTCACAGGGGCGTCTTGGCCCT 3775  
Db 768 GCGCAACGATGCGGCGCGCGGATGCGCGGGAACCGCCCTGAGAGGCGCACTTTGCCGG 827  
Qy 3776 CCAGCGGT 3783  
Db 828 CAGGGAT 835

RESULT 11  
US-09-252-991A-11911/c  
; Sequence 11911, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 11911  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11911

Query Match 3.6%; Score 155.4; DB 3; Length 879;  
Best Local Similarity 54.6%; Pred. No. 6.2e-18;  
Matches 387; Conservative 0; Mismatches 301; Indels 21; Gaps 3;  
Qy 3078 GTCTGCGAGGCGCAGAAAGCCACACGTCGGGATGTCTGGGTCTGCGGATCGGTT 3137  
Db 834 GTAGGCGGTCTCGGCGCGGCCAAGTAGGCGAAGAGTGGCAATTCTCTCGTTCGGGC 775  
Qy 3138 GAACGCGAGTTGCGCGCCAGTGTCCGCGCGCTCTGGGGGTCTGCGCGCGCGTTCGTT 3197  
Db 774 GTAGCGCGCAGCGCCATCAGGCGCTTGGGGCTCTGGGAGTCTCGCGTCTCGCGGTT 715  
Qy 3198 CATGTCTGCTTCCAGAAACCCGCGCGCACCGGTGACCGTGTATCCCGTTCGCCCGAG 3257  
Db 714 CATGTCTGATTCACCGGACCCGCGTGTGATGTTTACCGTGTATCCCGCGCGCGCGAG 655  
Qy 3258 TTGCTTGGCGGCGGCGGCTGAGCGTGTCCAGCGCACCTTGTGATCGCGTATCCGAT 3317  
Db 654 GTGCGGGCGGCGGCGGCTTGGTTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 595

Qy 3318 GGACTCGGGGACG---CGCGCCGGTTCGGCGACGAGATGTTGATGATCCGCGCCGCC 3374  
Db 594 GCGCGCGGAAACGCGATGCGTTTCGGCAATTGTTGCTCGGATGCTGATCACCAGCGCCGCC 535  
Qy 3375 GTCCGCGAGTCTGTTTTCAGTCCGTGCTGGACCAAGAAACGCGTTCGGACGCTTGACGCG 3434  
Db 534 GTCCGCGCATGTGTTGACCGCGCTTGGGTGGCGACGAAAGACGCTTGCGCACTTGATCGC 475  
Qy 3435 GACCAAGTCTGTCGAAGACCTCTCGGTGACTTTCGGTGTATCCGTCGGAGCGCTGACGCG 3494  
Db 474 CAGGCTCGCTCGAAGTTCGGCGCGGAGCTCTTCCACCGAAACGAGGCGGAGGACGCG 415  
Qy 3495 CGCGTGTTCACCAAGATGTCGAACCTCGGGCGGCACTCCGAACTCGCCCATCCCGCGCTC 3554  
Db 414 GCGGTGTGTGACCAAGATGTCGATGCG-----TCCGAAGCGAGCGCGCTTCGTC 364  
Qy 3555 GAAACCGCGTAGAGCGCGCGCGCTCACCCACGACGCGGAGTTCGGCCCGGATGCGCAA 3614  
Db 363 GACGGCGGCGACGAGCGCGCGCT-----CGGCGCTATCGGCATGACAGGCGCAG 313  
Qy 3615 CGCTGTTCGGCGCTGCTCGGATGCTTCGACGGTCTCTCGCGCGCGCGCTTCGCTGCT 3674  
Db 312 GATGCTGCTCGCGCGCTTTCGAGTTCGGCGCGCAGGCGCGGAGCTTGTCTCGCGAGCT 253  
Qy 3675 GCGGTAGTGGACTGCGACGAGCGCGCGCTTCGGCGCGCAGCGCGGATACCGCGTCC 3734  
Db 252 GACGTAGTGAAGCGACGCTGCGCGCTTCGGCGCGCAGCGCGGACGATGCGCGCGCC 193  
Qy 3735 GATGCGCGCGCTTCGCCCGCTCACAGGCGCGCTTCGCTTCGCTCCAGCGGT 3783  
Db 192 GATGCGCGGGAACCGCGCTTCGACGAGCGAGCTTTCGCGCGGAGGAT 144

RESULT 12  
US-09-489-039A-430/c  
; Sequence 430, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 430  
; LENGTH: 798  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-430

Query Match 3.5%; Score 150; DB 3; Length 798;  
Best Local Similarity 51.9%; Pred. No. 5.2e-17;  
Matches 389; Conservative 0; Mismatches 355; Indels 6; Gaps 2;  
Qy 3024 TCAGAAATCGTTTCGCGCTGTGGCATCGACGTACTGCGCGGTGATCCAGTGAGTCGTC 3083  
Db 798 TCAGATGCGCATTCCTCCCGGACACTTCAATGCGTGGGCAATTCACCGAGGATCATC 739  
Qy 3084 GAGGCGCAGAAAGCCACACGTCGCGGATGTCTGGGTCTGCGGATGCGGTTGAACAC 3143  
Db 738 GGATTAACAGCGGCGCATCATCTGCGCGATATCTCTGCGAGGCCAGGCGCGCGCGGC 679  
Qy 3144 GAGTGTGGCGCGCAGTCGCGCGCTTCGGGGTCTGCGG---CGCGGTGCGTTCAT 3200  
Db 678 CGTCATCGCAGCAGACGACCTGAGCGACCTGCGGGTTATCCCGCACACACCGCGCGTGA 619  
Qy 3201 GTCCGTCTCCAGAAACCGCGCGCCACCGCTGTGACCGTGTATCCCGGTTCCCGCAAGTTG 3260  
Db 618 ATCGTGGGATGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 559

3261 CTGGCCAGGCGAGCGTGTCTCCACCGCCCTTGGTCATCGGTATCCGATGGA 3320  
Db |||||  
558 AAACGCCATATAGCGGTTCAGGTTTCCACCGCGGATTCATGATCGATAGGGGCGCG 499  
Qy |||||  
3321 CTCGGGGAAACCGCGCGGTCGGGGCAGACGAGATGTGTGATGATCGCGCGCGCGTCCG 3380  
Db |||||  
498 GTTCGCCATCAGATGCGGGTGAGSCGGAGCGAAACATTCATATTGCGCGCGCATCGGC 439  
Qy |||||  
3381 CAGTCGTTTCACTGCTGTGACGACGAAACAGCGGTGCCCGGACGTTGACGGGACGAG 3440  
Db |||||  
438 CAGCAGCGGCAATATTTTGGCTGAGGAGAAACGCTCTTCACGTGTGATGCGATACAG 379  
Qy |||||  
3441 TCGGTGGAAGACCTCTCGGTGACTTCCGTGATCGTCCCGAGCGCTGACGCGCGGCTT 3500  
Db |||||  
378 CGCATCGAATTCGGTTCCGTTGGCGTTAAGAAATCCCATGCGCTGCGGCGCATT 319  
Qy |||||  
3501 GTTCACCAAGATGTGCAACTCGGGCGGCATCCGAACCTCGCCCATCCCGGCGTCAAGCGC 3560  
Db |||||  
318 ATTGACCAGATAATCAAAATTTATCG---CGCCAGCTCCGTGAGCGCGCTGAAAAGC 262  
Qy |||||  
3561 CGGTAGAGCGGGCGCGTCAACGACGACGCGGAGTTTCGGCCCGGATGCGCAAGCGCTG 3620  
Db |||||  
261 GCTAACAAATCGTCAAAAGTGTGCAATTCGCGCGCGCTGAAAGGCAACGCGATCGCCG 202  
Qy |||||  
3621 TCCGCGCTGCTCCGATGTTCTCGACGTTCTCTCGCGCGCGCTCGCTGCTGCGGTA 3680  
Db |||||  
201 GCGGCCAGTGCTTCAACCCGTTGTAACCTTCTGTTCCGCTCGGCAAGACGCGTTTGTGA 142  
Qy |||||  
3681 GTGACTGCCACGAGCGCCCGTTCGCGCGCCAGCGCGGCGGATACCGCGTCCGATGCC 3740  
Db |||||  
141 GGTCACTACTACATTTACGCGCGCTGTGCGAGCGCTTCGAGCGTTCGCGCGTCCAGGCC 82  
Qy |||||  
3741 CCGGTTCCCGGTCACAGCGCGGCTTT 3770  
Db |||||  
81 CCGGTCGCGCTGTTCACAGCGCGATTTT 52

## RESULT 13

US-08-125-468-1  
; Sequence 1, Application US/08125468  
; Patent No. 5589385  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, Michael J.  
; APPLICANT: Lotvin, Jason A.  
; APPLICANT: Strathy, Nancy  
; APPLICANT: Fantini, Susan E.  
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for  
; TITLE OF INVENTION: chlortetracycline and tetracycline Formation and cosmids  
; TITLE OF INVENTION: useful therein  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: American Cyanamid Company  
; STREET: One Cyanamid Plaza  
; CITY: Wayne  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07470  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/125,468  
; FILING DATE: 22-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Tsevdos, Estelle J  
; REGISTRATION NUMBER: 31,145  
; REFERENCE/DOCKET NUMBER: 31,255-02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201)831-3241  
; TELEFAX: (201)831-3305

; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30001 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-125-468-1

Query Match 3.4%; Score 145.6; DB 2; Length 30001;  
Best Local Similarity 53.9%; Pred. No. 3.8e-16;  
Matches 348; Conservative 0; Mismatches 289; Indels 9; Gaps 2;

Qy 3128 CGATGCGGTGAACACGAGTTGGCGGCCAGTGC CGCGCGCCCTCGGGGGTCTGCGGCC 3187  
Db 7734 CGAAGCGGAGACACGCTCTCTCGGTGTGCGCCAGGTGCGCGCGTACGCTGGCGGA 7793  
Qy 3188 GCCTGCGTTCACTCCGTCTCCAGAAACCGCGGCGACCGCTTGACCGTGATGCC 3247  
Db 7794 CCGCACGGCCATCGGCGTCTCGACGTAGCCGGGCGACGCGCTTGACCGTGTGCGG 7853  
Qy 3248 GTTCCCGCAGTTGCTGCGCAGGGCGAGCGTGTCCACCGCACCTTGTGTCATCG 3307  
Db 7854 TGTGGCGAGTTCTTGGCCAGCGCTTGTGTGAAGCGGATGACCGCGCTTGGAGGCG 7913  
Qy 3308 CGTATCGATGGAATCGCGGGAACCGCGCGCGGTCCGCGGACAGAGATGTTGATGCC 3367  
Db 7914 AGTAGGGGCGCCACGCGGACACCTGCTGTCGCGCGGTGAGCGCGATGATGCC 7973  
Qy 3368 GCCCGCGTCCGCGAGTGTGTTTCACTCGGTCTGCAACAGAAACAGCGGTGCCGAGCGT 3427  
Db 7974 GCCGTGCGCGCGCTCCATGCGCGCGGTGTGAGGACCTCGCGGTGACGCGGAAAG 8033  
Qy 3428 TGACGGGACGACGTGCGTCAAGACCTCTCGGTGACTTCGTTGATCGTCCCGAGCGC 3487  
Db 8034 CGCTGTGAGTTGTGTGATGATGCTCTCCACAGCTCTCGGTGAGGTGAGGTGA 8093  
Qy 3488 TGACGCGCGCTTGTTCACAGGATGTCGAACCTCGCGCGGCGACTCCGAACTCGCCCATCC 3547  
Db 8094 CGCCACCGCGTTGGGTCCGCGTGTGTA-----CCAGCAGCGCGATCGCGCGAAGC 8147  
Qy 3548 CGCGTCAACCGCGCGTGTGAGCGCGCGGTCAACCAACAGCGCGGATGTCGCGCCCGA 3607  
Db 8148 GGTCCACCGCGCGCGGACGAGCTGTACGTCCGCGCGGAGCGGACGTCGCGCGGA 8207  
Qy 3608 TGGCAACGCTGTCCGCGCTGTCCGAGTGTCTGACGCTCTCTCGCGCGCGCGCT 3667  
Db 8208 GCGGTTCACCTCCAGGCGCTCGCGCGCGGAGCGGTCTCTCGCGACGCGCTCGG 8267  
Qy 3668 CGCTGCTCGGTGTGACTGTCACAGCGCGCTCGCGCGCGGCGCGCGCGCGGCGATAC 3727  
Db 8268 CGGTGCGGCGCAGATGAG---ACGCTACGCGCGCGCGCGCGCGCTCGGCGCTGG 8324  
Qy 3728 CGCGTCCGATCCCGCGCTTCCCGGTTCCCGGTTCCACAGGCGGTCTTGCC 3773  
Db 8325 CCAGTCCGATCGCGTGTGTCGCGCGCGGTCAACAGGCGAGCGTGGCC 8370

## RESULT 14

US-08-474-933-1  
; Sequence 1, Application US/08474933  
; Patent No. 5866410  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, Michael J.  
; APPLICANT: Lotvin, Jason A.  
; APPLICANT: Strathy, Nancy  
; APPLICANT: Fantini, Susan E.  
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for  
; TITLE OF INVENTION: chlortetracycline and tetracycline Formation and cosmids  
; TITLE OF INVENTION: useful therein  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: American Cyanamid Company



STREET: One Cyanamid Plaza  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07470  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25-5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,933  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/125,468  
FILING DATE: 22-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsevdos, Estelle J  
REGISTRATION NUMBER: 31,145  
REFERENCE/DOCKET NUMBER: 31,255-02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201)831-3241  
TELEFAX: (201)831-3305  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30001 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-474-933-1

Query Match	3.4%;	Score 145.6;	DB 2;	Length 30001;
Best Local Similarity	53.9%;	Pred. No. 3.8e-16;		
Matches 348;	Conservative	0;	Mismatches 289;	Indels 9; Gaps 2;
3128	CGATGCGGTTGAACACGAGCTTCGGCGCCAGTGC	CGCGCGCGCGCTCGCGGGGTCTGCGGCC	3187	
Qy				
Db	CGAAGCGGAGAGCACTCTCTCTGGTGGTGC	CGCCAGGTTCGGGGGTACGCCTGGCGGA	7793	
Qy				
Db	GCGGTGCGTTTCATGTCGTCTCCACGAAACCCGGCGGCA	CCGGCGTTCACCGTATCCCCC	3247	
Qy				
Db	CCCGCACGGCCATCGGGCTCTCGACGTAGCCGGGGCAGAC	CGCGTTCACCGTGGTCCGG	7853	
Qy				
Db	GTTTCCCCAGTTGGCTGCCAGGGCGAGCGTAGCGTGTCCA	CGGCACCTTTGGTCAATCG	3307	
Qy				
Db	TGTGGGCGAGTCTCTTTGGCCAGCGCCTTGGTGAAGCCGAT	GAACCGCGCGCTTGGAGGCGG	7913	
Qy				
Db	CGTATCCGATGGACTCGGGGAAACGCGCGCGGGTCGGCGG	CAGACGAGATGTTGATGATCC	3367	
Qy				
Db	AGTAGGGGGCGCCAGCGGAGACCTTGTCTTGGCCCGTGG	AGGCGACGCTGATGATCC	7973	
Qy				
Db	GCCCGCGCTCGCGCAGTCTGTTTCAGTCCGTCTGGAC	CAGCGGTGCCCGGACGT	3427	
Qy				
Db	GCCCGTCCCGCGCGCTCCATGCGCGCGGTGGTAGGA	CCTCGCGGTGACGCGGAAGA	8033	
Qy				
Db	TGACGGGCAACAGTCGGTTCGAGACCTCTCGGTGACTT	CGGTGATCGGTCCCGAGCGCG	3487	
Qy				
Db	CGCTGGTGAAGTTGGTTCGATCAGTCTCTGCCACAGCT	CGTCGCTGAGGGTGGAGGTGA	8093	
Qy				
Db	TGACGCCCGGTTGTTTCCACGAGATGCGAACTCGGCGG	CACATCCGAACTTCGCCCATCC	3547	
Qy				
Db	CGCCACCGCCGTTGCTTCGGCGTGTGTGA-----	CCAGCAAGCCGATTCGGCGCGAAGC	8147	
Qy				
Db	CGGCGTCCGAACGCCCGCTAGACGCGGCCCGCTCAC	CCACGACGCGAGTTTCGCCCGCGGA	3607	
Qy				
Db	GGTCCACCGCGCCCGGACGAGCTGTCACTTCGGCGG	CGAGCGACGCTCGGCCCGGA	8207	
Qy				
Db	TGGCCAAACGCGTGTTCGCGCGGTCTCCGAGATG	GTCTTCGAGCGGTCTTCGCGCGCGCGCGCT	3657	
Qy				
Db	GGCGGTCCACTTCAGAGCCTTCGCGCGCAGCGGGG	GACCGTCTCGGCGACGCGCTTCGCG	8257	

Qy	3668	CGCTGCTGCCGTAGTGGACTGCACACGAGCCCGCGGCGCAGCCGCGAGGCGCATAC	3712
Db	8268	CGGTGCGGGGCGCAGATGAAG---ACGTTACAGCCCCGCGCGGCGCGCTTCGGCGCTGG	8324
Qy	3728	CGCGTCCGATGCCCCCGCGCTTCCC CGCTCAC CAGGGCGGTTC	3773
Db	8325	CCAGTCGATGCCGCTGGTCGCCCCGCTCAC CAGGGGCGACGTCGCC	8370

RESULT 15

```

US-09-902-540-336/c
; Sequence 336, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,893
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 336
; LENGTH: 1571
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-336

```

Query Match	3.2%	Score 139.6	DB 3	Length 1571
Best Local Similarity	53.0%	Pred. No. 3.3e-15		
Matches 410	Conservative 0	Mismatches 334	Indels 30	Gaps 4
QY	3037	CCGCGCTGTGGCATCGACGTACTGCGCGGTGATCCACCGTGTAGTCTGTGCGAGGCCAGAAAG	3096	
DB	1427	CCACCATCGACCGCGATAGAGTGCCTGTGTATGAATCCCGCGTGGCACATCGCCAGGTGC	1368	
QY	3097	GCCACGACGTCGGCGATGTCTGTGGGTCTGCCGATCGGTTTGAACGCGAGTTCGGCGCC	3156	
DB	1367	CCGACGGTGGCGCGATGTCTCGGGCTGTGCGGTAGTGTCTTAA-----GGGACATG	1317	
QY	3157	AGTCGCGCGCGCGCTCTGGGGGTCTGCGCGCGCGTGGTTCATGTCTCGTCTCCACGAAA	3216	
DB	1316	AGCGCCAACTCGGCGCGCGACGCGCGGCCCGGTCTCGCGATTTCATGTCTGTATCTCGTCAC	1255	
QY	3217	CCGCGCGCCACCGCGTTTGACCGTGTATCCCGCGTTCGCCAGTTCGCTGGCCAGGCGCAGC	3276	
DB	1256	CCGCGGTTCGATCAGCTTTCATGTGTATGCGCGCTCGCGACTTCCCGTGGCCAGGCGCTTC	1197	
QY	3277	GTGAGCGTGTCCACCGGACCCCTGGTTCATCGCGTATCCGATGGACTCG---GGGAAACGCG	3333	
DB	1196	GTGAGCGCGATCAGGCGGACTTGTCTCATCGAGTAGCGGTGACTCCGCCATAGGCGACG	1133	
QY	3334	CGCGGGTTCGGCGAGACGAGATGTTGATGATTCGCGCGCGCGTTCGCGAGTCTGTTTCACT	3399	
DB	1136	CGCTCGCGAAACGACGACCAATCGAGATGATGGCGGCCACCTTCGGCGATGTGCGCGAGC	1077	
QY	3394	CGGTCTGGACCGAGACAGCGGTGTCGCGGACGTTTGACGGCGACACGAGTCGGTTCGAAGACC	3455	
DB	1076	GCGGCTTGAGACGCGACGACACGCTCTCGAAGCTGGATTGCCAGGTCTCTGTGAGTTC	1011	
QY	3454	TCCTCGGTGACTTCCGTGATCCGTCCCGAGCGGTGACGCCCGCGGTGTTTTCACAGGATG	3513	
DB	1016	TCGAGTGTGACTTTCTTTCGAAACGGTCCGGAAGGGAAGATGCGCGGCGTTTTCACCAAGGATG	957	
QY	3514	TGGAACCTCGGCGGGCACTCCGAACTCGCCCATCCGGCGGTGGAACGCCCGGTAGAGCGCG	3577	
DB	956	TCCAGGCGG-----CCCAACTCGAGACCGCGCGGTTCGATCTGCGCCACGAGCGGCG	906	
QY	3574	GCGCGGTTCACCGACGACCGCGAGTTTCGCGCCGCGATGGGCCAACGCTGTCCGCGCTGTCTC	3633	

Db	905	TCGGGGTCTCCGTTC-----TCCGCCTGGATGGCCACGGCCCGTCGGCCGAGTGCC	855
Qy	3634	CGGATGGTCTCGACGGTCTCTCGCGCCGCCCGCTGCTGCTGGCCGTAGTGGACTGCCACG	3693
Db	854	TTGAGTTCAACCGACCCAGCGTGGCGCTTTGTCCGCCAGAGTTCCGCATAGGTGAAGGCTACG	795
Qy	3694	AGCGCCCGTCCGGCGCCAGCCGCGAGGGCGATACCGGTCGATGCCCGCGGCTTCCCGG	3753
Db	794	TCCGCCCTCCCTTTGCCAGATGCCGAACGATGGCGCGCGATGCCACACGCGTGCCACCT	735
Qy	3754	GTCACCGGGCGGCTCTTGCCCTCCAGCGGTCTTCCATACCTCGTCCCATGTGCA	3807
Db	734	GTCACCAACGGCGCTTGCCACTGAACGCCATCATGTCCACCTCTACTTTGGA	681

Search completed: January 19, 2006, 03:06:23  
Job time : 711 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2006, 09:46:57 ; Search time 15841 Seconds  
(without alignments)  
12836.100 Million cell updates/sec

Title: US-10-017-471b-19  
Perfect score: 4346  
Sequence: 1 gtcgacgacgcgtcggtt.....gcgtagagcgcgggtacc 4346

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_hic:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_est7:\*  
9: gb\_gsa1:\*  
10: gb\_gsa2:\*  
11: gb\_gsa3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query	Length	DB	ID	Description
1	150.6	3.5	837	9	CC620734	OGQAC91TM	CC620734 OGQAC91TM
2	150.6	3.5	870	9	CC629147	OGUAC220TC	CC629147 OGUAC220TC
3	148.4	3.4	917	10	CG096128	PUFZK93TD	CG096128 PUFZK93TD
4	146.8	3.4	864	10	CL963775	OBIFC038	CL963775 OBIFC038
5	145.4	3.3	582	6	CB854086	UI-CF-DU1	CB854086 UI-CF-DU1
6	136.8	3.1	1083	7	CK214283	FGA502620	CK214283 FGA502620
7	136.2	3.1	1257	10	CL948439	OBIFB005	CL948439 OBIFB005
8	135.6	3.1	843	7	CN823495	Oa_gplbn	CN823495 Oa_gplbn
9	135.4	3.1	800	7	CN145046	WOUNDI_26	CN145046 WOUNDI_26
10	134.2	3.1	825	10	CZ544259	SRAA-aad5	CZ544259 SRAA-aad5
11	134.2	3.1	890	8	DR887252	JGI_CUNC6	DR887252 JGI_CUNC6
12	133.8	3.1	764	7	CN130030	RHOH1_38	CN130030 RHOH1_38
13	132	3.0	780	7	CO202814	Oa_gplbn	CO202814 Oa_gplbn
14	132	3.0	809	9	CC398497	FUHLIC28TD	CC398497 FUHLIC28TD
15	131.6	3.0	795	9	BZ537000	OGAGU62TC	BZ537000 OGAGU62TC
16	131.2	3.0	845	8	DN258874	Meso03087	DN258874 Meso03087
17	130.6	3.0	2332	10	AG363333	Mus muscu	AG363333 Mus muscu
18	130	3.0	767	10	CG333497	OGIEE86TH	CG333497 OGIEE86TH
19	128.6	3.0	674	6	CD866271	AZO2.102P	CD866271 AZO2.102P
20	127.4	2.9	750	10	CL982611	OBIFC048	CL982611 OBIFC048
21	127.2	2.9	730	6	CA140180	SCEZRT202	CA140180 SCEZRT202
22	126.2	2.9	769	7	CO423399	GGEZHT100	CO423399 GGEZHT100

23	126.2	2.9	2332	10	AG363333	Mus muscu
24	125.8	2.9	798	10	CL974024	OBIFC042
25	125	2.9	696	6	CD866272	AZO2.102P
26	125	2.9	1081	7	CK211362	FGA502320
27	123.8	2.8	664	8	DN483106	root3_11
28	123.2	2.8	1885	10	AG363563	Mus muscu
29	122.6	2.8	697	6	CA130420	SCCRT100
30	121.8	2.8	782	6	CA258734	SCBGR7301
31	121.4	2.8	767	9	CG670424	OGWIY21TH
32	121	2.8	839	8	DR027220	bda010012
33	120.6	2.8	1569	10	AG341503	Mus muscu
34	118.8	2.7	841	10	CG096126	PUPZK93TB
35	118	2.7	709	1	AJ702939	AJ702939
36	117.2	2.7	795	10	CL958605	OBIFC001
37	117.2	2.7	828	9	CG670432	OGWIY21TV
38	117	2.7	861	6	CA104193	SCJFHR103
39	117	2.7	1569	10	AG341503	Mus muscu
40	116.6	2.7	778	6	CF637158	zmrw00_0
41	115.8	2.7	550	7	CN822865	Oa_gplbn
42	115.8	2.7	823	7	CN822937	Oa_gplbn
43	115	2.6	1821	10	CL090560	ISB1-17N1
44	114.8	2.6	461	8	CX949662	UMC-bcl_0
45	114.6	2.6	1821	10	CL090560	ISB1-17N1

ALIGNMENTS

RESULT 1  
LOCUS CC620734 837 bp DNA linear GSS 19-JUN-2003  
DEFINITION OGQAC91TM ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0364P13,  
genomic survey sequence.  
ACCESSION CC620734  
VERSION CC620734.1 GI:31987155  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 837)  
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Renick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
Consortium for Maize Genomics  
Unpublished (2002)  
TITLE Other\_GSSs: OGQAC91TC  
JOURNAL Contact: Cathy Whitelaw  
COMMENT TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TR  
Class: methylation filtered.  
Location/Qualifiers  
1. 837  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0364P13"  
/clone\_lib="ZM 0.7 1.5 KB"  
/note="Vector: pBCSK; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

ORIGIN  
Query Match 3.5%; Score 150.6; DB 9; Length 837;  
Best Local Similarity 52.3%; Pred. No. 4.6e-18;  
Matches 417; Conservative 0; Mismatches 359; Indels 21; Gaps 3;  
QY 3037 CGCCCTGTGGCATCGACGTACTGGCGGTGATCCACCGTGTGTCGGAGCCGAGAAAG 3096

Db 21 CCGCCGTTGACCTGATGACCTGCGCTTAAACCGAGCGCGGCTGCTGGCCAGGAAC 80  
Qy 3097 GCCACCAAGTCGGCGATGTGTCGTCGGGTCTGCGATGCGGTGGAACACGAGTTGGCGGCC 3156  
Db 81 GCCACCAAGTCGGCGATGTGTCGTCGGGTCTGCGATGCGGTGGAACACGAGTTGGCTCCACC 140  
Qy 3157 AGTCGCGCGCGCTCGGCGATGTGTCGTCGGGTCTGCGATGCGGTGGAACACGAGTTGGCTCCACC 3216  
Db 141 CTCGCGAGGAACGCTTCG-----TCCTTCCCGCCCAAGAACACAGCTCGGTGCGCAGCGGC 194  
Qy 3217 CCGCGCGCACCGGTTGACCGTGCATCCCGCTTCCCGAGTTCCTCGGCGCGAGCGGAGC 3276  
Db 195 CCGCGCGCACCGGTTGACCGTGCATCCCGCTTCCCGAGTTCCTCGGCGCGAGCGGAGC 254  
Qy 3277 GTGAGCGGTGTCCACCGACCTTGGTTCATGCGGTATCCGATGAGTTCGCGGAGCGCGC 3336  
Db 255 GTGATCGCTCCACCGCGCGGTGTTGTCGCGGTGACGCTGCGTATCCGCGAGGAGCGTG 314  
Qy 3337 CCGGTGCGCGGAGAGAGATGTTGATGATCGCGCGC-----GTGCGGAGTTCGTTTC 3390  
Db 315 CCAACGATGACGAGCAAGCTCACGATGCGCGCGCGCTGTTGGCTGGCAGCGGTTTC 374  
Qy 3391 AGTCGCTGCTGACGAGCAAGCGGTGCGCGAGCTTTCGCGGAGTTCGCGGAGTTCGGAAG 3450  
Db 375 GCCGCTCGCGGCGCACGAGGAACGTCCTCGCGGAGCTTCACCGGAGACATGCGGTGGAAG 434  
Qy 3451 ACCTCCTCGGTGACTTCGCTGATCCGTCGCCGAGCGCTGACGCGCGCGGTTGTTTACCAGG 3510  
Db 435 TCGTCGAGCGCGGTGTCGCGGAGCGCGGCTACTTGGCAATTGAGGAGCGCGCGCAGCAG 494  
Qy 3511 ATGTCGAATCGCGGCGCACCTCGGAATCGCGGCTTCGCGGAGTTCGAGCGCGGTAGAGC 3570  
Db 495 ACAACGATGTCGCGCGGCGCACCGAAGCTCTCTCGCGCGCGGTTCGAAGAGCGGTGACG 554  
Qy 3571 CCGCGCGGTGACCGACGAGCGGTGCGCGCGCGGTGCGCGGAGTTCGCGGAGTTCGCGGAG 3630  
Db 555 GCGTCGCGGTGCGGAGC-----TCGCGCGGAGCGCGGAGTTCGCGGAGTTCGCGGAG 605  
Qy 3631 CTCGCGATGTCGACGCGTCTCGCGCGCGCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 3690  
Db 606 GAGCGAGCTCGCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 665  
Qy 3691 AGGAGCGCGGTGCGCGGCG 3750  
Db 666 ACG 725  
Qy 3751 CCGGTACAGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3810  
Db 726 CCAGTGATGACGCGTACGCGCGCGGTGCGCGCGCGGTGCGCGCGCGGTGCGCGCGGTGCT 785  
Qy 3811 ATATCAGCGCGCGCGGT 3827  
Db 786 GCGACTGCGATGGCTGT 802

RESULT 2  
LOCUS CC629147/c  
DEFINITION OGUCR20TC\_ZM\_0.7\_1.5\_KB\_Zea\_mays\_genomic\_clone\_ZMMBMA0391D15,  
genomic survey sequence.  
ACCESSION CC629147  
VERSION CC629147.1 GI:32001217  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 870)  
REFERENCE Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
AUTHORS Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

TITLE Consortium for Maize Genomics  
JOURNAL Unpublished (2002)  
COMMENT Other\_GSSs: OGUCR20TM  
Contact: Cathy Whitelaw  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: IF  
Class: methylation filtered.  
Location/Qualifiers  
1. .870  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0391D15"  
/clone\_lib="ZM 0.7 1.5\_KB"  
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

FEATURES  
source  
1. .870

ORIGIN  
Query Match 3.5%; Score 150.6; DB 9; Length 870;  
Best Local Similarity 52.3%; Pred. No. 4.6e-18;  
Matches 417; Conservative 0; Mismatches 359; Indels 21; Gaps 3;  
Qy 3037 CCGCTGTGGCATGACGATGCTGCGGGTCTGCGGATGCTGCGGAGTTCGCGGAGC 3096  
Db 805 CCGCGTTGACCTGATGACCTGCGCGCTTAAACCAAGCCGCGGCTGCTGGCCAGGAAC 746  
Qy 3097 GCCACCAAGTCGGCGATGTCGCGGTCTGCGGATGCTGCGGAGTTCGCGGAGC 3156  
Db 745 GCCACCAAGTCGGCGATGTCGCGGTCTGCGGATGTCGCGGAGTTCGCGGAGC 686  
Qy 3157 AGTCGCGCGCGCTCGGCGGTCTGCGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCT 3216  
Db 685 CTCGCGAGGAACGCTTCG-----TCCTTCCCGCCCAAGAACACAGCTCGGTGCGCAGCGGC 632  
Qy 3217 CCGCGCGCGCGGTGACGCTGATCCCGGTTCGCCGAGTTCGCTGCGCGCGCGCGCGCG 3276  
Db 631 CCGCGCGCGCGCGGTGACGCTGATCCCGGTTCGCCGAGTTCGCTGCGCGCGCGCGCG 572  
Qy 3277 GTGAGCGGTGTCACCGACCTTGGTTCATCGCTGATCCGATGAGTTCGCGGAGCGCGCG 3336  
Db 571 GTATCGCTTCCAGCGCGGTGTCGCGGTGTCGCGGTGTCGCGGTGTCGCGGTGTCGCG 512  
Qy 3337 CCGGTGCGCGCGAGAGATGTTGATGATCCCGCGCG-----GTGCGCGAGTTCGTTTC 3390  
Db 511 CCAACGATGACGAGGAACGTCACGATGCGCGCGCGCGCGGTGTCGCGCGCGCGCGGTTC 452  
Qy 3391 AGTCGCTGTCGACGAGAACAGCGGTGTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCG 3450  
Db 451 CCGCGCTCGCGGCGACGAGGAACGTCGCGCGCGCGCGGTTCACCGGAGAACATGCGCGTCG 392  
Qy 3451 ACCTCCTCGGTGACTTCGCTGATCCGTCGCCGAGCGCTGACGCGCGGTGTCACCGAG 3510  
Db 391 TCGTCGACGCGCGGTGTCGCGGAGCGCGCGGTGTCGCGGAGTTCGCGGAGTTCGCGGAG 332  
Qy 3511 ATGTCGAATCGCGCGCGCATCTCCGAACTTCGCCATTCGCCGCGTTCGAAACCGCGGTGAGC 3570  
Db 331 ACAACGATGTCGCGCGCGGAGCCGAAACGCTCTCGCGCGCGGTTCGAAAGAGCGGTGACG 272  
Qy 3571 CCGCGCGGTGACCGACGAGTTCGCGCGCGGTTCGCGGAGTTCGCGGAGTTCGCGGAG 3630  
Db 271 GCGTCGCGGTTCGACACG-----TCGCGCGCGGAGCGCGCGCTGTCGCGCGCGG 221  
Qy 3631 CTCGCGATGCTGACGCGTCTCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3690  
Db 220 GAGGCGAGCTCGCGCGCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 161  
Qy 3691 ACAGCGCGCGGTGTCG 3750

Consortium for Maize Genomics  
Unpublished (2002)  
Other\_GSSs: OGUCR20TM  
Contact: Cathy Whitelaw  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: IF  
Class: methylation filtered.  
Location/Qualifiers  
1. .870  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0391D15"  
/clone\_lib="ZM 0.7 1.5\_KB"  
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

```

Db      160  ACGCGCGCGCCAGGTGGCGAGTTGGGAGGACACCTCGCGCGCGATGCGCGGAGCGG 101
QY      3751  CCGGTGACAGAGCGCGTCTTCCCTCCAGCGTCTTCATACCTCGTCCCATGTGACGCG 3810
Db      100  CCAGTGATCAGCGCTACGCGCGCGTTCGAGCGGAGCAGCAGCGTGTCTTCGCGCGTT 41
QY      3811  ATATCAGCGCGCGCGT 3827
Db      40  GCGACTGCCATGGCTGT 24

RESULT 3
LOCUS   CG096128                917 bp    DNA        linear    GSS 20-AUG-2003
DEFINITION   PUFZK93TD ZM 0.6-1.0 KB Zea mays genomic clone ZMMBTA0768017,
              genomic survey sequence.
ACCESSION   CG096128
VERSION     CG096128.1  GI:33978422
KEYWORDS    GSS.
SOURCE      Zea mays
            Zea mays
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 917)
AUTHORS    Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
            Maize Genomics Consortium
            Unpublished (2003)
            Other_GSSs: PUFZK93TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TP
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source           1..917
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTA0768017"
                     /clone_lib="ZM 0.6-1.0 KB"
                     /note="Vector: pCR4-TOPO; Site:1: EcoRI; 0.6-1.0 kb high
                     Cor selected genomic DNA library"

ORIGIN
Query Match      3.4%; Score 148.4; DB 10; Length 917;
Best Local Similarity 54.3%; Pred. No. 1.2e-17;
Matches 408; Conservative 0; Mismatches 311; Indels 33; Gaps 4;

QY      3037  CCGCTGTGGCATCGACGTACTGGCGGTGATCCACCGTAGTCGTCGAGGCGCAGAAAG 3096
Db      192  CCGCGGTGGCGCGATGACCTGGCGGTGACCACTCGCGCGCGTCGTCGACAGAAAG 251
QY      3097  GCCACACGTGGCGATGTCGTGGGTCTGCCGATGCGGTGACACGAGTTGGCGGC 3156
Db      252  CCGACACCGCGCGATGTACCGGCTCGCGA---GCTCTGCATCGGTTGTCTCC 308
QY      3157  AGTGGCGCGCGCTCGCGGGTCTCGCGCGCGTCTATGTCCTCTCCACGAAA 3216
Db      309  ACGGCTGTGCGACCATGGCTCGCTCTTGCCCTTGAAGAACATGTCGTCGCGTGGCC 368
QY      3217  CCGGGCGCCACCGCGTGTGACCGGTATCCCGCGTTCGCCAGTGTGCTGCCAGGCGAGC 3276
Db      369  CCGGGCGCGACGCGAGTTGGCGGTGACGCGGTGCCCGCGAGCTCTTGGCCATGTGTCGC 428
QY      3277  GTGAGCGTGTCCACCGCACCTTGTGTTGTCGCGTATCCGATGACATCGGGGAACGCGCGC 3336

```

```

Db      429  ACCAGCGCTCCACCGCGCTTGGACGCGGTGTAGGCCCGAACCGCGCGGAGGCTC 488
QY      3337  CCGGTCCGCGCAGACGAGATGTTGATGATCCGCGCGCGTGC-----GGCAGTGGTTTC 3390
Db      489  CCCACACGAGCAGATGTCA CGCGCAGATGCGCCCGCGCGCGCGCGCGGAGGCGGTTG 548
QY      3391  AGTCCGTGCTGACAGCAAGCAAGCGGTGCCCGAGAGTTGACCGCGCAGCAGTGGTTCGAAG 3450
Db      549  GCGGCTCCGAGGAGCAGAGNACGCGCGCGCAGAGTTCA CGCGAGCAGCGGTTCGAAG 608
QY      3451  ACCTCTCGTGACTTCGCTGATCCGTCGAGCGCGTGA CGCCCGCGTGTTCACGAGG 3510
Db      609  GCC-----TCGTTGGCGGTGCGCGCACCGCGGGGTAGCTGTGCTCCAGCACC 656
QY      3511  ATGTGAACTCGGCGGCACTCCGAACTCGCCCATCCCGGCGTTCGNAACGCGCGTAGAGC 3570
Db      657  CCGGCGCTGGCCACAGGATGTGCGCGCGCGAAGGCTCTGTCGCGCGCGTCGGAAGAGC 716
QY      3571  GCGGCGCGGTCAACCCACGACGCGGAGTTGCGGCCCGAGTGGCCAAAGCCTGTCCGCGCTG 3630
Db      717  GACCGCGCGCGGCTCTGTCGAGACGTCCGCGCACGCGCCACCGC-----G 764
QY      3631  TTCCGATGCTCGACCGGTCTCTCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3690
Db      765  CGCGGAGCGACGCGCGCGCAGCGCGCGCTCTGTCGCGCGCGCGCGCGTAGCCAGCACC 824
QY      3691  ACGACGCGCGCTCGCGCGCGCAGCGCGCGAGGCGATACCGGTCGCGTCCGATGCCCGCTTCCC 3750
Db      825  AGGCTCGCGCGCGAGGAGGACAGGTGCGTGGCGGNTGGCGCGCGCGATGCCCGCGCGCG 884
QY      3751  CCGGTCCACGAGGCGGTCTTGGCCCTCCAGCGG 3782
Db      885  CCGGTACGATCGCACGCGCGCGCGCGCGCGG 916

RESULT 4
LOCUS   CL963775/c                864 bp    DNA        linear    GSS 21-SEP-2004
DEFINITION   OsIFC038648 Oryza sativa Express Library Oryza sativa (indica
              cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL963775
VERSION     CL963775.1  GI:52382266
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
            Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
            1 (bases 1 to 864)
            Ma,L., Wang,C., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
            An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
            Unpublished (2004)
            Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
            Location/Qualifiers
     source           1..864
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "

ORIGIN

```

Query Match	3.4%;	Score 146.8;	DB 10;	Length 864;
Best Local Similarity	52.6%;	Pred. No. 2.4e-17;		
Matches 400;	Conservative 0;	Mismatches 342;	Indels 18;	Gaps 3;
Qy	3037	CCGCTCTGGCATCGAGTACTCGCGTGTATCCACCGTGTAGTCTCGGAGGCCAGAAAG	3096	
Db	854	CCGCCATTGACCAATGACTTGGCCGTGTATCCACCGCGGGGTCTGCTGCAGAGGAAG	795	
Qy	3097	GCCACCAAGTCGGCGATGTCTGCGGTCTGCCGATGCGGTGTGAACACGAGATTGGCGGCC	3156	
Db	794	CCACCAACCGCGCCAGCTCATCGGTCTGCCGATCGGCCCATCGGCGCACTCGCTCTGCG	735	
Qy	3157	AGTCCCGCGCGCTCGGGGTCTGCGCGCGTGGTGTTCATGTCCGCTCCACGAAA	3216	
Db	734	ACGCGACGACGCGCTCTCTGCTCT---TGCGCGGTAGAACATCGCGGTCTGCGACCGCG	678	
Qy	3217	CCCGCGCCACCGGTTGACCGTGTATCCCGCTTCCCGATTCCTTGGCCAGCGCGAGC	3276	
Db	677	CCCGCGCCACGCTGTTCCGCGTATCCCGTTCGCGGGAGCTCTTGGCCAGCACCTTC	618	
Qy	3277	GTGAGCGTGTCCACCGCACCTTGGTATCGGTATCCGATGATCGATGATCGGGGAAACGCGGC	3336	
Db	617	GTATCGCTCTCACCGCGCTTCTGTCGACATGACGCGCGTACCGCGGCGCGAGGAGC	558	
Qy	3337	CGGTCGCGGACGAGATGTTGATGATCGCGCGCGCTGCGC-----GCAGTCTGTTTC	3390	
Db	557	CCCAAGTTGAGACGAGAGAGTGAACACGCGCGCGCGCGCGCGCGCGCGCGCGCG	498	
Qy	3391	AGTCGCTGTGACACGAGAACAGCGGTGCGCGGACGTTTGACGCGGACACGATCGGTGGAAG	3450	
Db	497	GCGCGCTCGCGGACGACAGAAAGTGCAGCGCGCGGTTTACGGGAGACGCGGTCCAC	438	
Qy	3451	ACCTCTCTGGTGTGATTCGGTATCCGTCGAGCGGTGATCGGCGCGGTGTTTCCACGAG	3510	
Db	437	TGCTCCGCGACGATGTCCGACGCGCGGTGACGCGCGTCTTGCACGCGCGCGCGCG	378	
Qy	3511	ATGTCGAACTCGGCGGACATCCGAACTCGGCGCATCCCGGGTTCGAGCGCGGTAGGC	3570	
Db	377	ACCAGGACGTTCAGGTTCGCGCGCGGAAACGCTCTTGCAGCGCGGTGGAACAGCGCG	318	
Qy	3571	GCGCGCGGTCAACACGACGCGGATTCGCGCGCGATGCGGCGGATGCGGCGCGCTG	3630	
Db	317	TGCGCGGTTCGACACGTCGCGCGCACGCGGACGCGCGCGCGCGCGCGCGCGCG	258	
Qy	3631	CTCCGATGTGTTCT-----CGAGGTCTCTCGCGCGCGCGCTCTGCTCTCGGTAG	3681	
Db	257	GTGCGGAGGCGTTGAGCGCGCGACGAGCTTCTCGCGGATCTGGGTCTCGCGACGTAG	198	
Qy	3682	TGGACTGCAAGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	3741	
Db	197	CCGACCAACGACG	138	
Qy	3742	CGGCTTCCCGCGTCAACGAGCGCGTCTCTGCGCTCCAGCG	3781	
Db	137	CGCGCGCGCGCGTACGATAGCTACCGCGCGGTGGAAGCG	98	

RESULT 5  
CB854086/c  
LOCUS  
DEFINITION  
UI-CF-DUI-aak-p-04-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone  
CB854086  
CB854086.1 GI:30044463  
EST.  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 582)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene

discovery  
Genome Res. 6 (9), 791-806 (1996)  
8889548  
Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 FORWARD  
POLYA=No.  
FEATURES  
source  
1..582  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-DUI-aak-p-04-0-UI"  
/tissue\_type="Primary Lung Epithelial Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-DUI"  
/notes="Organ: Lung; Vector: p7T3-Pac (Pharmacia) with a modified polylinker; Site 1: Ecor I; Site 2: Not I; UI-CF-DUI is a normalized cDNA library containing the following tissue(s): Primary Lung Epithelial Cells The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into p7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GGCTGTAGGC.  
TAG\_SEQ=None found"

Query Match	3.3%;	Score 145.4;	DB 6;	Length 582;
Best Local Similarity	55.0%;	Pred. No. 4.5e-17;		
Matches 306;	Conservative 0;	Mismatches 247;	Indels 3;	Gaps 1;
Qy	3047	CATCGAGTACTGCGCGGTGATCCACCGTGTGATCGGAGGCCAGAAAGGCCACCGT	3106	
Db	582	CTTCGATACGTCGCGGTGATCCAGTTGTCGCATCGCCAGCAGGCTCGAATTCAC	523	
Qy	3107	CGGCGATGTCTCGGGGTCTGCCGATGCGGTGTGAACACGAGAGTTGGCGGCCAGTCCGCGG	3166	
Db	522	CCCCAATATCTTCGCGCAACCGCGCACGCGCGGTGTTGTCGCCACCATCGCAT	463	
Qy	3167	CCGCTCTCGGGGTCTGCCCGCGC---CGTGGTTCATGTCGCTCTCCAGAAACCGCGG	3223	
Db	462	TCACCTGTGGTATTTCGCGCACTCGCGCGCGCTGGAAGTCGGTGGCAATCGCGCGGGTG	403	
Qy	3224	CCACCGGTTGACCGTGTATCCCGTTCCTCCCGTTCGCTCGCCAGCGCGAGCGGTGAGCG	3283	
Db	402	CCAGGTTTTCAGCGTGTATGCCAGCGCGCGCGAGTCTTTGGCTCTGATAACCGACAGCA	343	
Qy	3284	TGTCACCGCACCTTGGTCTATCCGATATCCGATGAGTCTCGGGAAACCGCGCGCGGTTCG	3343	
Db	342	CTTCCACGCGCGCTTTCATCGACGATAGGCGGCGTAGCGCGCGAGGCTGAACCGCGCA	283	
Qy	3344	CGGACAGCAGATGTTGATGATCGCGCGCGTTCGCGCAGTCTGTTTCAGTCCGTCTCGA	3403	
Db	282	GCGCGTGGAGATGTTGATGATGCGCGCGCGCATCGCTGATCAGCGGCGACGAGTTTTTGG	223	

3404 CCAGGAACAGCGGTGCCCGGACGTTGACGGCGACAGTCGCTCGGAGACCTCTCGGTGA 3463  
|||||  
222 TGAGGAAACACGCGCCCTTTGAAGTGAATGGCCACCACTTTGTGCAACTGCTCTTCGGTGG 163  
|||||  
3464 CTTCCGTGATCGTCCCGAGCGCGTACGCCCGCGTGTGTTCACCAAGGATGTGCAACTCGG 3523  
|||||  
162 TTTCCGCAAGCTTGGTGAGCGCGCATGCGCGCTTGTGTATCAGGAATTAATATT 103  
|||||  
3524 GCGGCACCTCCGAACCTGCCCCATCCCGGGGTGCAACCGCGGTAGAGCGCGCGCGGTAC 3583  
|||||  
102 GCTGGCGGAAACGCTTTTGAGCACCGTAGCGATTGTGCCCGACAAAGCATCGAAGTTGC 43  
|||||  
3584 CCACGACGCGGAGTTC 3599  
|||||  
42 CGCTCTGGCTGACGTC 27  
|||||

RESULT 6  
CK214283/c 1083 bp mRNA linear EST 09-DEC-2003  
LOCUS FGAS026206 Triticum aestivum FGAS: Library 6 CAP GATE 1 Triticum  
DEFINITION aestivum cDNA, mRNA sequence.  
ACCESSION CK214283  
VERSION CK214283.1 GI:39620387  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum

Spematophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 1083)  
Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
Linka, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,  
Penniket, C., Roach, J.L. and Sarhan, F.  
Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
Unpublished (2003)  
Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033

Email: fgas\_estcs.usask.ca  
This sequence is the direct result of the Base calling software  
Phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [66,721].  
Plate: L6B012 row: F column: 03.

Location/Qualifiers  
1. .1083  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum FGAS: Library 6 CAP GATE 1"  
/note="Organ: Crown and leaf; Vector: pCMV.SPORT6; Crown  
(50%) and leaf (50%) tissues from wheat cultivar Norstar  
after short exposure times to low temperature in the light  
and in the dark. 12 mRNA populations were combined before  
constructing the library. The first 6 populations: After 7  
days of growth at 20°Cs from wheat cultivar Norstar after  
short exposure times to low temperature in the light and  
in the dark. 12 mRNA populations were combined before  
constructing the library. The first 6 populations: After 7  
days of growth at 20, wheat plants were transferred to 4C  
in the light. 1cm crown sections and green leaf tissue were  
separately harvested after 1, 3, and 6 hours of low  
temperature exposure. The last 6 populations: After 7 days  
of growth at 20C, wheat plants were transferred to 4C in  
the dark. 1cm crown sections and green leaf tissue were  
separately harvested after 1, 3, and 6 hours of low

temperature exposure. First strand synthesis in this  
library was done in the presence of methylated dCTP  
thereby protecting from internal cleavage with NotI. In  
addition, this library used a primer for second strand  
synthesis that annealed to an artificial sequence (RNA  
oligo) added before first strand synthesis. Therefore when  
sequences from EST generated from this library will be  
masked for vector and adaptor sequences, an additional  
masking step will have to be included to mask this RNA  
oligo that is common to all clones (sequence  
CGACTGGAGCAGGACACTGACATGACTGAAGGAGTAGAAA)."

ORIGIN

Query Match 3.1%; Score 136.8; DB 7; Length 1083;  
Best Local Similarity 52.8%; Pred. No. 1.9e-15;  
Matches 391; Conservative 0; Mismatches 33; Indels 12; Gaps 4;  
QY 3032 TCGTTCCCGCTGTGGCATCGACGTACTGGCCCGGTGATCCACCGTGAAGTCGTGCGAGGCCA 3091  
DB 943 TTGTGCCCCCGTTGAGGCCCGACACTGGGCCCTGACCCACCCCGCGGTCTTTGCCGA 884  
QY 3092 GAAAGGCCACACGCTGCGCGATGTCTGGGTCTGCCGATGCGGTGAAACAGGAGTTGG 3151  
DB 883 GGAACCCCAAGACCGCGCGATGTCTGGCATGCCCCGAGCGGCCCCACCGTCGCGCTCG 824  
QY 3152 CGGCCAGTGGCGCGCGCGCTCGGGGTCTGCGCGCGCGCGTGTCTATGTCTCGTCTCCA 3211  
DB 823 CGATGTACCGCGCGCGCTCTCTCATCCGCTTTTCCCGTTG--TAGTACATCGGGGTGGCG 766  
QY 3212 CGAAACCCCGCGCGCACCGGTGTGACCGTGTATCCCCCGCTTCCCGCAGTTTCCCTGGCCAGGG 3271  
DB 765 TGGACCCCGCGCGCACCGGTGTGCGGTGTATCCCGTGAACCGCGAGCTCTTTGGCCAGGA 706  
QY 3272 CGAGCGTGAAGTGTCTCACCGCACCTTGGTTCATCGGTATCCGATGATCGGGAACG 3331  
DB 705 TCCTGTAATGCTCTCCACCGCGCGCTTGTGACCGCGTACGCGCGGTAGCGCGCGCGCA 646  
QY 3332 CGCGCGGGTTCGCGCGACGAGAGTGTGATGAT-----CGCGCGCGCTCGCGCAGTC 3385  
DB 645 GCGACCCGACCCCGCGACGAGAGGTGACATGCGGCGCGCGCGCTCGCGGTGAGCC 586  
QY 3386 GTTTCAGTCCGTGTGGACAGGAAACAGCGGTGCGCGGAGTGTGACGCGGAGCCAGTCCGT 3445  
DB 585 GTTTCGCGCTTCCCGAGGACAGGAGGTGCGCGCGGTGTGACACCGGAGCGCGCT 526  
QY 3446 CGAAGACCTCTCTCGGTGATCTCCGTGATCCGTCCCGAGCGCTGACGCGCGCGGTTCCTCA 3505  
DB 525 CGTACGTGCGCTCTGCTCTCGCGCAGCGCGGTGTACGAGTAGTCCAGCACCGCGCGCTG 466  
QY 3506 CCAGGATGTGAATCTCGGCGCGGAC--TCCGAATCTGCCCATCCGCGGTTCGAACGCGCG 3564  
DB 465 TCGTGACGAGCACGTGTCAGCTCCCATCCGAAAGCGCGCGCGCTCGAACAGCGCC 406  
QY 3565 TAGAGCGCGCGCTGTCACCCAGCGCGAGTTCGCGCGCGATGCGCAACGCTCTCCG 3624  
DB 405 CGCACATGCGCGCTGCGACACGCTCCG---CTCCACCGCGATGATGACGCTCGCG 349  
QY 3625 CGCGTCTCCGATGATCTCGACGCTCTCTCGCGCGCGCGCTCTGCTGTGCGGTAGTGG 3684  
DB 348 TGGCGCGCTTGTATCCCGCCACGACCTCTTTGGCGCGCGCGCGCGCGCTAGGCC 289  
QY 3685 ACTGCCACAGCGCGCGCTCGCGCGGACGCGGAGGCGGATACCGCGTTCGATGCCCGG 3744  
DB 288 ACCGCGCACGCGCGCGCGCGAGGCTGTGAGTGTGTTGACACGCGCGAGCGGATGCCCG 229  
QY 3745 CTTCCCGCGTCCACGAGGC 3764  
DB 228 GCACCGCGCTCAGCATCGC 209

RESULT 7  
CL948439/c 1257 bp DNA linear GSS 21-SEP-2004  
LOCUS CL948439



```
DEFINITION OSIFS005157 Oryza sativa Express Library Oryza sativa (indica
cultiivar-group) genomic, genomic survey sequence.
ACCESSION CL948439
VERSION CL948439.1 GI:52360448
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 1257)
AUTHORS Ma L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G.K.S., Deng, X.W. and Wang, J.
TITLE An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
source
1..1257
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences"
ORIGIN
Query Match 3.1%; Score 136.2; DB 10; Length 1257;
Best Local Similarity 53.1%; Pred. No. 2.4e-15;
Matches 313; Conservative 0; Mismatches 273; Indels 3; Gaps 1;
QY 3193 GCGTTTCATGTCGCTCTCCACGAAACCCGGCGCCAGCGTTGACGTCGATCCCGCGTTCC 3252
DB 590 GCGTGGATCTGGTATCGATCAGGCGGGCGCGCAGCGGTTTCACGCGGATGCCCTCCGCT 531
QY 3253 CCCAGTTCCCTGGCAGGGCGAGGTGAGCGGTGTCACCGCACCCCTTGGTCATCGCGTAT 3312
DB 530 GCCACTTCGCGCGAGGCCACCACCGTAGACAGTCGATGGCGGCCCTTGGCTGCCGCATAG 471
QY 3313 CCGATGGACTCGGGAAACGGCGC--GCCGGGTGCGGGCAGACGAGATGTTGATGATCGCG 3369
DB 470 TCCACGTACTGGTGGCGCGCCCGCCAGCGGGCGCGCGCTGGAACATTGACGATGCTG 411
QY 3370 CGCGCGTCGCGCAGTCGTTTCAGTCCGTGCTGGACAGGAAACAGCGGTGCCCGGACGTTG 3429
DB 410 CGCCTCTGGCGCGCGCTGAGGCTCATGCGCCGCAAGGCTCGCGCGCAGTAGAAG 351
QY 3430 ACGGCGACAGTCGCTCGAAGACCTCTCTCGGTGACTTCGTTGATCCGTCGCGAGCCGCTG 3489
DB 350 CTGCCCAGACATTGATGTGGAACATGCGCCGACGCGCTGGCCGCTCATCTCGGACAGC 291
QY 3490 AGCCCGCGTGTTCACACAGATGTGCAACTCGGGCGGCATCTCCGAACTCGCCCATCCG 3549
DB 290 CGTGAGGACAGTCCTCACACCGCGCATTTGTTACACAGGGCGGACAGCGCGGCCCATTCG 231
QY 3550 GCGTCGAACCGCGCTAGAGCGCGCGCGCTCACCCACAGCCCGAGTTTCGGCCCGGATG 3609
DB 230 GTGTCACCGCTTCGAACAGCCGCAAGACTCTGCGACTGTGCGCCACGTCGCGCTGGATG 171
QY 3610 GCCAAGCGCTGTCGCGCGCTGCTCCGGATGCTTCAGACAGGTCTCTCGCGCGCGCGCTCG 3669
DB 170 GCCACGGCTTCGCCGCTTCGCGCCCGGATGGCATCGACACGCGCGCGCGCGCGCG 111
QY 3670 CTGCTGCGGTAGTGAATGCGACAGAGCGCCCGCTCGCGGCGCAGCGCGAGCGGATACCG 3729
110 TCGCGCGCATAGTTTCACTCCACGGCCAGCGCGCGCGCCAGCGCGCGTGGCG 51
QY 3730 GGTCCGATGCCCGCGCTTCCCGCGGTACACAGGGCGGTCTTGCCCTCA 3778
DB 50 GCGCGGATGCCCGCGCTCCCTCCGTCACCAAAACCACTTTTGTGCA 2
RESULT 8
CN823495 843 bp mRNA linear EST 02-JUN-2004
LOCUS Oa splbn_05P05_M13reverse Sheep spleen\brain pSport1 library Ovis
DEFINITION aries cDNA clone Oa_splbn_05P05 5', mRNA sequence.
ACCESSION CN823495
VERSION CN823495.1 GI:47951564
KEYWORDS EST.
SOURCE Ovis aries (sheep)
ORGANISM Ovis aries
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Caprinae; Ovis.
REFERENCE 1 (bases 1 to 843)
AUTHORS Gossner, A. and Hopkins, J.
TITLE Ovine spleen\brain cDNA library
JOURNAL Unpublished (2004)
COMMENT Contact: J Hopkins
Veterinary Biomedical Sciences
University of Edinburgh
Summerhall Square, Edinburgh, EH9 1QH.
Email: j.hopkins@ed.ac.uk
Plate: 05 row: P column: 05
Seq primer: M13reverse
High quality sequence start: 16
High quality sequence stop: 547.
FEATURES
source
1..843
/organism="Ovis aries"
/mol_type="mRNA"
/db_xref="taxon:9940"
/clone="Oa_splbn_05P05"
/clone_lib="Sheep spleen\brain pSport1 library"
/note="Vector: pSport1"
ORIGIN
Query Match 3.1%; Score 135.6; DB 7; Length 843;
Best Local Similarity 53.7%; Pred. No. 3.2e-15;
Matches 388; Conservative 0; Mismatches 305; Indels 30; Gaps 4;
QY 3060 GCCGTGATCCACCGTGAAGTTCGGAGGCGCAGAAAGCCACCGTCGGCGATGTCGTC 3119
DB 109 GCCGTGATATAGCGGCTCAGGGCTTACCAGGTAGCGAGAACTGGCGATCTCTC 168
QY 3120 GGGTTCGCGATGCGGTTGAACAACGAGTTCGGGCGCAGTGGCGGCGCGCTCGGGGT 3179
DB 169 CACATGGCCATACCGGCCACGGCCATGAAGCCGATCAG-----GCTCTCGCGAA 219
QY 3180 CTGCGCGCGCGTGGTTCATGTCGTCCTCCACGAAACCGCGCGCCACCGGTTGACCGT 3239
DB 220 ATCACCACCTTCGCGGTTTCATGTCGTTGTCCACCGGCGCTCGCTGACGTTATTGATG 279
QY 3240 GATCCCGCGTTCCCGCAGTTCCTGGCCAGCGCGCGAGCGTGAAGCGTGTCCACCGCACCCCTT 3299
DB 280 GATGCGACGTGGCCCGCAGTTCGCGGCCACAGGCTTTGTGAGGCCGACGAGCGCGACTT 339
QY 3300 GGTTCATCGCGTATCCGATGGACTCGGGGAACG---CGCGCGGGTTCGGCGGACGAGAT 3356
DB 340 GCTCATGCGGTACGCGGCGCCACCGGCAACCGGCAATGCGATCGGCAATGGTGTCTCCGAT 399
QY 3357 GTTATGATCGCGCGCGCTCGCGAGTCGTTTTCAGTCCGTCGTGGACGAAACAGCGG 3416
DB 400 GTTATGATGCGGCGCACCTTCACCCATGTGCGCGCGCTTCCTGGTGGGACAAACAC 459
QY 3417 TGCCCGGACGTTGACGCGGCGACGAGTCGTCGAGACCTCTCTCGGTGACTTCCTCGTATCCG 3476
```

Db	460	GCTCGCAGCTGTGATGCGAGGGGTACCGGTGGAAGTCTTCCAGGCTGAAATCTTCCAGCGG	519
Qy	3477	TCCCGAGCGCTGACGCGCCGCGTGTGTTTAC CAGGATGTGAAATCGGGCGGCAC TCCGAA	3536
Db	520	GCCGACGGCCAGCAGCGCGGGTGTGTTCCAGGATATCCA-----GGCGGCCGGA	570
Qy	3537	CTCGCCCATCCCGCGGTGCBAAACGCCGCTAGAGCGGGCCGCTCACCCACGACGCGCGAG	3596
Db	571	GGCTTCGACGGTGGCTGTGACCGCGCTGCGGATGGCGCGGCAT-----CGGCGCT	621
Qy	3597	TTCGGCCCGGATGGCGCAACGCTGTCCGCGCGCTCTCCGGATGGTCTCGACGGTCTCTCG	3656
Db	622	GTCCGCGATGGATCGCCACGGGCTTTGCGGCTCTTCGCTGATCAGCTGTCTGTGCAATGCTTC	681
Qy	3657	CGCCGCGCCTCTCGCTGTCTGCGTAGTGGACTGCCACGAGCGCCCGCTCCGCGGCAGCCG	3716
Db	682	GGCTTTGGCGGCCAGCTGACATAGTGTGAGGCGACGGCTGCGCCCTGNGGGGCCAGGCG	741
Qy	3717	CAGGGCGATACCGCGTCCGATGCCCCCGCTTCCCGGTCACGAGGCGGTCCTTGCCCTC	3776
Db	742	CTTGACGATGGCGCGCCGATACCGCGGACCCGCCCTGGATCAGGCGCACTTTGCCGCTG	801
Qy	3777	CAG	3779
Db	802	AAG	804

RESULT 9	
CN145046/c	
LOCUS	800 bp mRNA linear EST 01-APR-2004
DEFINITION	WOUND1_26_B03.g1_A002 Wounded leaves Sorghum bicolor cDNA clone
ACCESSION	WOUND1_26_B03_A002 5', mRNA sequence.
VERSION	CN145046
KEYWORDS	
SOURCE	CN145046.1 GI:45983920
ORGANISM	EST. 1
	Sorghum bicolor (sorghum)
	Sorghum bicolor
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
	clade; Panicoideae; Andropogoneae; Sorghum.
	1 (bases 1 to 800)
	Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C.,
	Sun,F., Sullivan,R., Brady,J., Eastman,A., Miller,V., Gonzalez,M.,
	Amuso,C., Chhabra,D., Johnson,H., Kamran,D. and Pratt,L.H.
TITLE	A Sorghum EST database: mechanically damaged and methyl
	jasmonate-treated leaves
JOURNAL	unpublished (2003)
COMMENT	Other ESTs: WOUND1_26_B03.b2_A002
	Contact: Cordonnier-Pratt MM
	Laboratory for Genomics and Bioinformatics
	The University of Georgia, Department of Plant Biology
	Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
	Tel: 706 542 1860
	Fax: 706 583 0210
	Email: mmpratt@uga.edu
	Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
	the Human Genome Center, University of Tokyo Institute of Medical
	Science; plant material and RNA prepared at Texas A & M University;
	sequencing done in the Laboratory for Genomics and Bioinformatics,
	University of Georgia. Sequence ends have been trimmed to exclude
	vector and regions below Phred quality 16. Three-prime sequences
	are presented as their reverse complement and have been trimmed to
	exclude polyA.
FEATURES	Seq primer: Sug5 (CTTCTGCTCTAAAGCTGG).
	Location/Qualifiers
	1..800
	/organism="Sorghum bicolor"
	/mol_type="mRNA"
	/cultivar="BTx623"
	/db_xref="taxon:4558"
	/clone="WOUND1_26_B03_A002"
	/lab host="DH10B-T1 phage-resistant E. coli"
	/clone lib="Wounded leaves"

/note="Organ: Leaf; Vector: pME18S-PL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA<sup>+</sup> RNA harvested from 8-day-old hydroponically grown, Brx523 sorghum seedlings. For some plants, one-half of the second leaf was crushed without damaging the midvein. For others, methyl jasmonate was added to the growth medium to a final concentration of 100  $\mu$ M. Leaves were harvested 3 and 27 hr after treatment and pooled. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-PL3 vector (5'-prime DraIII site is CACTGTGTG, 3'-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

## ORIGIN

	Query Match	3.1%	Score 135.4	DB 7	Length 800
	Best Local Similarity	54.7%	Pred. No. 3.4e-15		
	Matches 321	Conservative	0	Mismatches 251	Indels 15
	Gaps				
Qy	3202	TCGCTGCCACGAAACCGCGCGCACCGCTTGACACCGTGATCCCGCTTCCCGCAGTTGC	3261		
Db	695	TCGGTGGCACGGGCCCGCGCGACACGTTTCGCACTCACTCCCTTCGCGCCACCTCC	636		
Qy	3262	CTGGCCAGGCGGAGCGTAGCGGTGCACCGCACCTTTGGTGCATCGCGTATCCGATGCAC	3321		
Db	635	TTGGCTAGGATCTTCGTGCATCGCTCCACGCGCGGTGGTCGCGGTGTACGCGCGTAT	576		
Qy	3322	TCGGGGNAACGCGCGCGGTTCGGGACAGAGATGTTGATGATCGCGCCCGCC-----G	3375		
Db	575	CCCGCGAGGAGCGTGCAGCATGAGACGACGAGAACTCACATGATCGCGCCCGCTGTTG	516		
Qy	3376	TCGCGCAGTCGTTTCAGTCCGTCGTGACAGGAAACACGCGTGCCTCGGACGTTGACGCG	3435		
Db	515	GCTGGGACTCGTTTCGCGGCTCACGGCACACGAGGAACGTCCTCGCGCACGTTCA	456		
Qy	3436	ACCACTCGGTGGAAGACCTCTCTCGGTGACTTCGTGTGATTCGTCCTCGAGCGCGTGA	3495		
Db	455	ACCATGCGTTCGAAGTCTTCGACGCGCGGTGTCCGCGAGCGCGCGGTACTTGTGCTGAGG	396		
Qy	3496	GCGTTGTTCCACGAGATGTCGAACTCGGGGGGCACTCCGAACCTCGCCATCCCGCGCTCG	3555		
Db	395	ACGCTTCACAGCAGACGACGATGTGCGCGGGACCCGGAACGCTCTCGCGCGGTGCG	336		
Qy	3556	AACCGCGGTAGACGCGCGCGGTCAACCAACGACGCGGATTCGCGCCCGGATGCCAAC	3619		
Db	335	AAGAGCGCGCACGCGCTCCGGTCTGACACG-----TCCGCGCGGACGCGCCACG	285		
Qy	3616	GCTGTTCGCGCGCTGCTTCGAGATGGTCTTCGACGCTCTCTCGCGCGCGCGCTCGCTGCTG	3675		
Db	284	GCCTGGTGGCGCGGGAGGAGCTCCGCGACGAGCTGTTCTGCTTCGCGGGAATTGAC	225		
Qy	3676	CGGTAGTGGACTGCGACGAGCGCCCCCTTCGCGGCCACCGCGAGGGGGGATACCGCGTTCG	3735		
Db	224	CGGTAGTTACACGACGCGCGGCCCGGAGGGCGCGAGGTGGGAGGACACCTCGCGGCG	165		
Qy	3736	ATGCCCGCGCTTCCCGGTGCACAGGGCGGTCTTGGCTCCAGCGG	3782		
Db	164	ATGCGGGGAGCGCGCGGTGATGAGGCGCACGCGGCGGTGAGCGG	118		

## RESULT 10

RESOLU I  
CZ544259

LOCUS

### DEFINITION

ACCESSION

**VERSION**  
**KEYWORDS**

**KEYWORDS**  
**SOURCE**

**SOURCE:**  
**ORGANIZATION:**

## ORGANIZATION

## REFERENCES

CZ544259 825 bp DNA linear GSS 13-MAY-2005  
 SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library  
 (SRAAGS 004) Strongyloides ratti genomic, genomic survey sequence  
 CZ544259  
 CZ544259.1 GI:64674801  
 GSS.  
 Strongyloides ratti  
 Strongyloides ratti  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Pangorhinoidea; Strongyloidea; Strongyloidea.  
 1 (bases 1 to 825)  
 Mitreva, M., McCarter, J.P., Thompson, F., Viney, M., Pape, D.,

Ritter,E., Martin,J., Wylie,T., Dante,M., Waterston,R.H., Clifton,S.W. and Wilson,R.  
Genome Survey sequences from the rat parasitic nematode Strongyloides ratti  
Unpublished (2005)  
Contact: Mitreva M  
Washington University in St. Louis  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: nematode@watson.wustl.edu  
Genomic DNA was provided by Fiona Thompson (F.Thompson@bristol.ac.uk) and Mark Viney (Mark.Viney@bristol.ac.uk) at the University of Bristol, Bristol, UK.

Class: shotgun.  
Location/Qualifiers  
1. .825  
/organism="Strongyloides ratti"  
/mol\_type="genomic DNA"  
/strain="isofemale line ED321 heterogonic"  
/db\_xref="taxon:34506"  
/dev\_stage="infective larval stage (il3)"  
/lab\_host="GS10"  
/clone\_lib="Strongyloides ratti whole genome shotgun library (SPRAGSS 004)"  
/note="Vector: pORW13; Site 1: BstX1; Site 2: BstX1; Strongyloides ratti genomic DNA was randomly sheared, end-repaired and size fractionated to enrich for 2-4 kb fragments. Genomic DNA was provided by Fiona Thompson (F.Thompson@bristol.ac.uk) and Mark Viney (Mark.Viney@bristol.ac.uk) at the University of Bristol, Bristol, UK. Sequencing by Washington University Genome Sequencing Center, St. Louis, MO."

ORIGIN  
Query Match 3.1%; Score 134.2; DB 10; Length 825;  
Best Local Similarity 56.0%; Pred. No. 5.8e-15;  
Matches 275; Conservative 0; Mismatches 213; Indels 3; Gaps 1;

Qy 3037 CCGCCTGTGGCATCGACGTACTGCGCGGTGATCCACCGTGTGTCGTGGAGCCAGAAAG 3096  
Db |||||  
Qy 3097 GCCACCACTGCGCGATGTCGTGGGTCTGCGGATGCGGTTGAACACGAGTTGCGGCC 3156  
Db |||||  
Qy 301 GAAACCAACGAGGCGATGTCCTCGGCTGACCCAGGCGCTGCAGCGCGCGGCGCTTCGCA 360  
Db |||||  
Qy 3157 AGTGC CGCGCGCGCTCGGGGTCTGCGCGCGCTGCGTTCATGTTCGTCTCCACGAA 3216  
Db |||||  
Qy 361 AAGTTCGGATCTGCTCTCGGTCTTGGCGCGAGGAAGCG---CCGTGGCGATCGGG 417  
Db |||||  
Qy 3217 CCGCGCGCACCGCGTTGACCGTGTATCCCGCTTCCCGCAGTTGCTGCCAGGCGCAGC 3276  
Db |||||  
Qy 418 CCGCGCGCACCGCGTTTACCGTGTATGTCGGCGCGCGAGTCTCTGCGAGACGTGTC 477  
Db |||||  
Qy 3277 GTGAGCGTGTCCACCGCACCTTGTGTATCGGTTATCGATGGAATCGGGGAAACCGCGCC 3336  
Db |||||  
Qy 478 GTGAAGGCTTCACCGCGCGCTTGGTGGCGTTGTAGACCGCATAGCCCGCGAGGTTTCAGC 537  
Db |||||  
Qy 3337 CGGTCGGCGCAGACGATGTTGATGATCGCGCGCGCTCGCGAGTCGTTTCAGTCCG 3396  
Db |||||  
Qy 538 GCGAGCACCGTGCTCGAAGATTTGACGATGCGCCCGCCCTCGTTGAGACGCCCGCGGCC 597  
Qy 3397 TGTGAGCAGGAAACAGCGGTGTCGCGGAGCTTGAACGGGACCGACGTCGTCGAGACCTCC 3456  
Db |||||  
Qy 598 TCGCGACGGTGTGAAGTTCGCGCACGTTGATGTGGAAGTCTGTCGTAGAGCGTG 657  
Qy 3457 TCGGTGACTTCGTGATCCGTCGCCAGCGCTGACGCGCGCGTGTGTTACACGAGTATCG 3516  
Db |||||  
Qy 658 TCGGTGTGTCGCGCAGCGGCGGCTTCTTGAACACGCGCGCGGTATTATTCACGACGTCG 717  
Db |||||

Qy 3517 AACTCGGCGCG 3527  
Db |||||  
Db 718 ACCTTGCCAG 728

RESULT 11  
DR887252  
LOCUS  
DEFINITION  
DR887252 890 bp mRNA linear EST 01-AUG-2005  
IMAGE:7796552 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Xenopus tropicalis (western clawed frog)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.  
REFERENCE  
1 (bases 1 to 890)  
Richardson,P., Lucas,S., Rokhsar,D., Detter,J.C., Ng,D.C.,  
Brokstein,P. and Lindquist,E.A.  
DOE Joint Genome Institute Xenopus tropicalis EST project  
Unpublished (2004)  
Other ESTs: JGI\_CUNC695.rev  
Contact: Lindquist,E.A., Richardson,P.  
DOE Joint Genome Institute  
2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
Tel: 925 296 5600  
Fax: 925 296 5710  
Email: cdna@jgi-psf.org  
Tissue Procurement: Dan Buchholz (Yun-Bo Shi Laboratory, NIH)  
cDNA Library Preparation: DOE Joint Genome Institute:  
http://www.jgi.doe.gov  
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov  
Clone Distribution: I.M.A.G.E. Consortium/LINL:  
http://image.llnl.gov  
Naming Conventions: EST name is generated by the concatenation of  
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'  
indicates a forward sequencing read of the insert. It does not  
necessarily reflect the orientation of the insert.  
Plate: CUNC 0005 row: n column: 6  
High quality sequence stop: 852.  
Location/Qualifiers  
1. .890  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="IMAGE:7796552"  
/tissue\_type="whole embryo"  
/dev\_stage="Metamorphic (st.64)"  
/lab\_host="Electromax DH10B"  
/clone\_lib="NIH\_XGC\_tropMet4"  
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This  
library was made from dr primed cDNA and cloned into  
Invitrogen pCMVSPORT6 vector. The work was done at DOE  
Joint Genome Institute. Poly A RNA were primed with 5'  
GACTAGTCTTAGATCGCGAG CGGCGCGCTTTTCTTTT 3'. CDNA  
were ligated to SalI adapter (5' TCGACCCACGGTCCG and  
5'CGACGCGTGG), digested with NotI, size fractionated in  
1.1% agarose gel electrophoresis and ligated into NotI and  
SalI digested pSPORT1 vector."

ORIGIN  
Query Match 3.1%; Score 134.2; DB 8; Length 890;  
Best Local Similarity 53.7%; Pred. No. 5.8e-15;  
Matches 386; Conservative 0; Mismatches 303; Indels 30; Gaps 4;

Qy 3060 GCCGTGTATCCACCGTGTAGTCTGTCGAGGCCAGAAAGCCACCGTCGCGGATCTCGTC 3119  
Db |||||  
Qy 94 CCGGTGTATATAGCGGCTCAGGCGCTACAGGCTACGAGTAGCGACGAACTGGCATCTTC 153  
Db |||||  
Qy 3120 CGGTCTGCCGATGCGGTTGAACACGAGTTGGCGGCCAGTCCGCGCGCGCTCGGGGT 3179  
Db |||||

Db 154 CACATGGCCATACCGGCCCCAGGCCCATGAAGCGGATCAG-----GCTCTCGCGGNA 204

Qy 3180 CTGCGCGCCCGTGGCTTCATGTCGTCTCCACGAACCGCGCCACCGCTTACCGT 3239

Db 205 ATCACCACCTTGGCGGGTTTCATGTCGTGTCACCGGGCTGGCTGCACGTATTATGATGT 264

Qy 3240 GATCCCGCGTTCGCCAGTTGCTGCGCCAGGGGAGCGTGAGGGTGTCCACGCCACCTT 3299

Db 265 GATGCCACCTGGCCCGCGAGTTCGCGGCCACGGCTTTGGTTCAGGCCACCGCGCGACTT 324

Qy 3300 GGTCTATCGCGTATCCGATGCTGCGGGAACG--CGCGCGGGTTCGCGGCAGACGAGAT 3356

Db 325 GCTCATGGCGTACGGGCCCGCCACCGGCCAACGGCATCGCATGCTGCTGCTGCGGAT 384

Qy 3357 GTTGATGATCCCGCGCGCTCGCGCATGCTGTTTCATGTCGTGCTGGACAGGAACAGCGG 3416

Db 385 GTTGATGATGGCGCACCTTCAACCATGTCGCGCGGCTTCTTGGTGGCGGACAAACAC 444

Qy 3417 TCGCGGAGCTTGACGGCGACAGTCGCTGCGGAAGACCTCTCGTGACTTCGTGATCGG 3476

Db 445 GCTGCGCAGCTTGATGGCGAGGTACGGTTCGAAGTCTTCAGGCTGAATCTTCCAGCGG 504

Qy 3477 TCCCGAGCGCTGACGCGCGTGTGTTTACACAGGATGTCGAATCTCGGCGGCACTCCGAA 3536

Db 505 GCGACGCGCAGCACCGCGGCTGTGTTTACACAGATATCA-----GGGCGCCGAA 555

Qy 3537 CTCGCCATCCCGGCTCGAAGCCCGCTAGAGCGCGCGCTCACCCAGCAGCGCGAG 3596

Db 556 GGTTCGACGGTGGCTTACGCGGCTGCGGATGGCGGCGCAT-----CGGCGCT 606

Qy 3597 TTCGCGCGGATGGCCAGCCCTGTCCGCGCTGCTCCGGATGGTCTCAGCGTCTCTCG 3656

Db 607 GTCGGATGATCGCCAGGCGCTTTCGCGCTTTCGCTGATCAGCGTCTTCTGCAATGCTTC 666

Qy 3657 CGCGCGCGCTCGCTGCTCGCTAGTGGACTGCGACGAGCGCGCTCGCGCGCGCGG 3716

Db 667 GGCCTTGGCGCGCGAGCTGACATAGTGAAGGCGACGGCTGCGCTTCGCGCGCGCGG 726

Qy 3717 CAGGGCGATACCGCGTCCGATGCGCGGCTTCCCGCGCTCACCGAGCGCGTCTTGCCT 3775

Db 727 CTTGAGATGGCGCGCGCATACCGCGGACCGCGCTGATCAGGCGACTTTCGCGCT 785

## RESULT 12

CN130030/c  
LOCUS  
DEFINITION  
RHOH1\_38\_H01.g1\_A002 Acid- and alkaline-treated roots Sorghum  
bicolor cDNA clone RHOH1\_38\_H01\_A002 5', mRNA sequence.

ACCESSION  
CN130030

VERSION  
CN130030.1

KEYWORDS  
GI:45958210

SOURCE  
Sorghum bicolor (sorghum)

## ORGANISM

Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

## REFERENCE

## AUTHORS

1 (bases 1 to 764)  
Cordonnier-Pratt M.-M., Suzuki Y., Sugano S., Klein R.R., Liang C.,  
Sun P., Sullivan R., Lim S., Eastman A. and Pratt L.H.

## TITLE

## JOURNAL

## COMMENT

Unpublished (2003)  
Other ESTs: RHOH1\_38\_H01.b1\_A002  
Contact: Cordonnier-Pratt MM  
The University for Genomics and Bioinformatics  
Laboratory for Genomics and Bioinformatics  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
the Human Genome Center, University of Tokyo Institute of Medical  
Science; plant material and RNA prepared at Texas A & M University;  
sequencing done in the Laboratory for Genomics and Bioinformatics,

University of Georgia. Sequence ends have been trimmed to exclude  
vector and regions below Phred quality 16. Three-prime sequences  
are presented as their reverse complement and have been trimmed to  
exclude polyA.

Seq primer: Sug5 (CTTCTGCTCTAAAAGCTGG).

## FEATURES

## source

1..764

/organism="Sorghum bicolor"

/mol\_type="mRNA"

/cultivar="BTx623"

/db\_xref="taxon:4558"

/clone="RHOH1\_38\_H01\_A002"

/lab\_host="DH10B-T1 phage-resistant E. coli"

/lab\_lib="Acid- and alkaline-treated roots"

/note="Organ: Root; Vector: pME18S-FL3; Site\_1: XhoI;

Site\_2: XhoI; The library was prepared from polyA+ RNA

from 8-day-old roots harvested from BTx623 sorghum

seedlings grown in hydroponic culture. HCl was added to a

pH of 3.0 to some seedlings, KOH to a pH of 9.0 for

others. Roots were harvested 3, 12 and 27 hr after

addition and pooled for RNA preparation. Double-stranded

cDNA was cloned unidirectionally into different DraIII

sites of the pME18S-FL3 vector (5-prime DraIII site is

CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises

the cDNA insert."

## ORIGIN

Query Match 3.1%; Score 133.8; DB 7; Length 764;  
Best Local Similarity 54.5%; Pred. No. 6.9e-15;  
Matches 320; Conservative 0; Mismatches 252; Indels 15; Gaps 2;

Qy 3202 TCGGTCTCACGAACCGCGGCCACCGGTGACCGTGATCCCGTTCCTCCCACTTGC 3261

Db 696 TCGGTGCGCACGGGCGCGCGCGACACGTTGCGAGTCACCTCTTCGCGGCACCTCC 637

Qy 3262 CTGGCCAGGCGGAGGTGAGCGTGTCCACGCAACCTTGGTCACTCCGATCCGATGAGAC 3321

Db 636 TTGGCTAGGATCTTGTCTATCGCTTCCAGCGGGGTGGTGGCGGTGTACGCGCGTAT 577

Qy 3322 TCGGGGAACGCGCGCGGTTCGCGGACGAGATTTGATGATCCGCGCGC-----G 3375

Db 576 CCGCGGAGGCGTTCGACGATGACGACGAGAACGTACGATGCGCGCGCGCTGTTCG 517

Qy 3376 TCGCGCAGTCTGTTTCAGTCCGTGTCGACGAGAACGAGGTGCGCGGAGCTTACGCGCG 3435

Db 516 GCTGGGACTCGGTTCGCGCCTCACGGCACACGAGGAACGTCCCGCGCACGTTTACCGCG 457

Qy 3436 ACCAGTCTGTCGAGACCTCTCTCGGTGACTTCGCTGATCCGTCGCGGCGGCTGACGCGCC 3495

Db 456 AACATGGCGTGAAGTCTCGACGCGGTGTCGCGAGGCGCGGGTACTTGTGCTGAGG 397

Qy 3496 GCGTTGTTTACAGGATGTGAACTCGGGCGGCATTCGGAATTCGCCATTCGCGGGGTGCG 3555

Db 396 ACGCTTGCACAGCAGACGATGTGCGGCGGGGACCCGAAACCTCTTCGCGGCGGTGCG 337

Qy 3556 AACGCGCGCTAGACGCGCGCGGTTCACCGACGCGCGGATTCGGCGCGGATGGCCAC 3615

Db 336 AAGAGCGCGCACGCGGTCTCGGCTCTGACG-----TCCGCGGAGCGGCGACG 286

Qy 3616 GCGTGTCCCGCGTCTCGGATGCTCGAGGTCCTCGCGGCTCTCTCGCGCGCGCGCTCGCTG 3675

Db 285 GCTGTGTCGCGGAGGCGGAGCTCCGACGAGTCTGCTGCTCTGCGGGAATTGGAC 226

Qy 3676 CCGTAGTGACTGCCACGAGCGCGCGTTCGCGGCGAGCGCGGAGGCGGATACCGGTCCG 3735

Db 225 GCGTAGTTTCCACGACGCGCGCGCGGCGGAGGCGGAGGTGGAGGACACCTCGCGCGCG 166

Qy 3736 ATGCCCGCGTTCCTCCCGGTTCACGAGGCGGTCTTCCCTCCAGCGG 3782

Db 165 ATGCCCGGAGCGCGCGGTGATGAGCGGACGCGCGCGCTCGAGCGG 119

## RESULT 13

```
CO202814      780 bp      mRNA      linear      EST 21-JUN-2004
LOCUS      Oa_splbn_14N14_M13_Reverse_Sheep_spleen\brain pSport1 library Ovis
DEFINITION      aries cDNA clone Oa_splbn_14N14 5', mRNA sequence.
ACCESSION      CO202814
VERSION      CO202814.1  GI:49013989
KEYWORDS      EST.
SOURCE      Ovis aries (sheep)
ORGANISM      Ovis aries
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Caprinae; Ovis.
REFERENCE      1 (bases 1 to 780)
AUTHORS      Gosner,A. and Hopkins,J.
TITLE      Ovine spleen\brain cDNA library
JOURNAL      Unpublished (2004)
COMMENT      Contact: J Hopkins
Veterinary Biomedical Sciences
University of Edinburgh
Summerhall Square, Edinburgh, EH9 1QH.
Email: j.hopkins@ed.ac.uk
Plate: 14  row: N  column: 14
Seq primer: M13 Reverse
High quality sequence start: 15
High quality sequence stop: 546.
FEATURES             source
     source
1..780
     /organism="Ovis aries"
     /mol_type="mRNA"
     /db_xref="taxon:9940"
     /clone="Oa_splbn_14N14"
     /clone_lib="Sheep_spleen\brain pSport1 library"
     /note="Vector: pSport1"

ORIGIN
Query Match      3.0%; Score 132; DB 7; Length 780;
Best Local Similarity 54.0%; Pred. No. 1.5e-14;
Matches 376; Conservative 0; Mismatches 290; Indels 30; Gaps 4;

QY 3060 GCCGTGATCCACCGTAGTCGTGGAGGCCAGAAAGCCACACGTCGGGAGTCTGTC 3119
DB 109 GCCGTGATATAGCGGCTCAGGCGCTACAGGCTAGGAGGACGAACTGGCGATCTTTC 168
QY 3120 GGGTCGCGATGCGGTGAACACGAGTTGGCGCCAGTGC CGCGCGCGCTCGGGGT 3179
DB 169 CACATGGCCATACCGGCCACCGCCATGAAGCCGATCAG-----GCTCTCGCGAA 219
QY 3180 CTGCGCGCGCGGTGGTTCATGTCGTCTCCACGAAACCGCGCGCCACCGGTTGACCGT 3239
DB 220 ATCACACTTGGCGGTTTCATGTCGTCTCCACCGGCGCTGGCTGCAGCTTATTGATGGT 279
QY 3240 GATCCCGCGTTCCTCCAGTTGCTCGGCCAGGCGAGCGTAGCGTGTCCACGCACTTT 3299
DB 280 GATGCCAGTGGCCCGCAGTTCGCGGCCAGGCGCTTTGGTCAGGCCGACAGCGCGACTT 339
QY 3300 GGTATTCGGTATCCGATGGACTTCGGGACG---CGCGCGGTCGGGCGAGACGAGAT 3356
DB 340 GCTCATGCGTACGGCGCGCCACCGCGCAACCGGCATCGCATGGCATTTGGTCTGCCGAT 399
QY 3357 GTTATGATTCGCCCGCGCGTTCGCGAGTCTGTTTCAGTCCGTCTGCAACAGGACAGGG 3416
DB 400 GTTATGATGCGGCGACCTTCACTCCATGTCGCGCGGCTTCTGGTGGCGACAAACAC 459
QY 3417 TGCCCGGAGTTGACGGGACAGTCGTCGAGAACCTCTCGTGGTGAATTCGTCGATCCG 3476
DB 460 GCTGCGCAGTTGATGGGAGGAGTACGGTTCGAAGTCTTCCAGGCTGAAATCTTCCAGCG 519
QY 3477 TCCGAGCGCGTCAGCCCGCGTTGTTACACAGGATGTCGAACTCGGGGGGCACTCCGAA 3536
DB 520 GCCGACGGCCACGACCGCGCGTGTGTTACACGAGATATCCA-----GGCGCCCGAA 570
QY 3537 CTCGCCCATCCCGCGGTTCGAACCGCGGTAGAGCGCGCGCGGTTCACCCACGACCGCAG 3596
1.
```

```
DB 571 GGCTTCGACCGTGGCGTTGACGGCGCTGCGGATGCGCGCGGCAT-----CGGCGCT 621
QY 3597 TTCCGCCCGGATGCCAACGCGCTGTCGCGCGCTGTCGCGATGCTCTCGACGGTCTCTCG 3656
DB 622 GTCGGCATGGATCGCCAGGGCTTTGCGCGCTTCGCTGATCAGCTGTGTTCTGCAATGCTTC 681
QY 3657 CGCGCGCGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3716
DB 682 GGCCTTGGCGCGCGCTGATAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 741
QY 3717 CAGGCGCATACCGCGTCGATGCGATGCGCGCGCTTCCTCC 3752
DB 742 CTTGACGATGCGCGCGCGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 777

RESULT 14
CC398497/c
LOCUS      CC398497 809 bp      DNA      linear      GSS 19-MAY-2003
DEFINITION      PUHLC28TD ZM_0.6_1.0 KB Zea mays genomic clone ZMMBTa476B07,
genomic survey sequence.
ACCESSION      CC398497
VERSION      CC398497.1  GI:30878587
KEYWORDS      GSS.
SOURCE      Zea mays
ORGANISM      Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE      1 (bases 1 to 809)
AUTHORS      Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
TITLE      Maize Genomics Consortium
JOURNAL      Unpublished (2003)
COMMENT      Other_GSSs: PUHLC28TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES             Location/Qualifiers
     source
1..809
     /organism="Zea mays"
     /mol_type="genomic DNA"
     /strain="B73"
     /db_xref="taxon:4577"
     /clone="ZMMBTa476B07"
     /clone_lib="ZM_0.6_1.0 KB"
     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
     Cot selected genomic DNA library"

ORIGIN
Query Match      3.0%; Score 132; DB 9; Length 809;
Best Local Similarity 52.9%; Pred. No. 1.5e-14;
Matches 307; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

QY 3202 TCCGTCTCCAGGAAACCGCGCGCCACCGGTGACCGTTCCTCCCGGTTCCCGGAGTTGC 3261
DB 762 TCCGTGCGCACAGGCGCGCGCCACACAGTTGGCGGTGATCCCTTCGCGCCACTCCC 703
QY 3262 CTGGCAGGGGAGCGGTGAGGTGTCCACCGCACCTTGGTTCATGCGGTATCCGATGCAC 3321
DB 702 TTGGCAGGACCTCGTTCATCGCTTCCACAGCGGCGTTGGAACGCGCGCTAG 643
QY 3322 TCGGGGAACGCGCGCGGTGCGGCGACAGAGATGTTGATGATCCGCGCGCGCTCGCG 3381
DB 642 CCGCGCGGACGACCGGCAACCGTCTGTGGACGAGAACGTCATGATGCGCGCCCGCTTTG 583
QY 3382 AGTCCTTTTCAGTTCGCTGCTGGAGCCAGGAACAGCGGTGCCCGACGTTGACGCGCA 3441
1.
```



This page blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2006, 09:42:01 ; Search time 2374 Seconds  
(without alignments)  
12200.823 Million cell updates/sec

Title: US-10-017-471B-19  
Perfect score: 4346  
Sequence: 1 gtcgacgacgcgcgcgggtt.....gcgtagagcgcgcgggtacc 4346

Scoring table: IDENTITY\_NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_21.\*  
1: Geneseqn1990s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*  
14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description	
		Match	Length				
1	4346	100.0	4346	6	ABN86379	Abn86379 ScbA, Scb	
2	4346	100.0	4346	10	ADE14794	Ade14794 Streptomy	
3	401	9.2	401	6	ABN86377	Abn86377 S. coelic	
C	401	9.2	401	6	ABN86378	Abn86378 S. coelic	
5	401	9.2	401	10	ADE14790	Ade14790 Streptomy	
6	316.2	7.3	19016	8	AAD47222	Aad47222 Streptoco	
7	274.2	6.3	88400	13	ADU47667	Adu47667 S. cyaneo	
C	255.4	5.9	7584	2	AAX19362	Aax19362 Rhodococc	
C	255.4	5.9	7600	3	AAAS1878	Aaas1878 Rhodococc	
10	239	5.5	4721	14	AEA81294	Aea81294 Streptomy	
C	11	238	5.5	936	13	ADW38421	Adw38421 DNA seque
12	229	5.3	699	6	AAL44719	Aal44719 S. virgini	
C	13	219.8	5.1	14638	14	ACL64643	ACL64643 M. xanthu
C	14	216	5.0	744	14	ACL65960	ACL65960 M. xanthu
C	15	213	4.9	741	13	ADT44601	Adt44601 Bacterial
C	16	210.6	4.8	762	13	ADS55812	Ads55812 Bacterial
C	17	210.6	4.8	762	13	ADS50097	Ads50097 Bacterial
C	18	206.6	4.8	750	13	ADS62269	Ads62269 Bacterial
C	19	206.6	4.8	750	13	ADS62467	Ads62467 Bacterial

C	20	206.6	4.8	750	13	ADS59685	Bacterial
C	21	206.6	4.8	750	13	ADS62885	Bacterial
C	22	196.6	4.5	24379	2	AAT93095	Streptomy
C	23	196.6	4.5	24379	2	AAV25925	Streptomy
C	24	194.2	4.5	687	14	AEA81295	Streptomy
C	25	185.4	4.3	732	13	ADS59899	Bacterial
C	26	184.6	4.2	912	11	ABD00980	Klebsiell
C	27	176.6	4.1	777	13	ADS59919	Bacterial
C	28	175.4	4.0	8169	2	AAV26609	Actinomad
C	29	174.6	4.0	762	13	ADS61934	Bacterial
C	30	172.8	4.0	765	13	ADS63580	Bacterial
C	31	172.8	4.0	765	13	ADS63953	Bacterial
C	32	172.6	4.0	747	13	ADS59965	Bacterial
C	33	171.4	3.9	747	13	ADS60268	Bacterial
C	34	169.2	3.9	753	13	ADT41762	Bacterial
C	35	167.6	3.9	732	13	ADS64319	Bacterial
C	36	164	3.8	726	13	ADS55246	Bacterial
C	37	162.8	3.7	720	13	ADS49731	Bacterial
C	38	162.2	3.7	786	8	ADA70196	Rice gene
C	39	162	3.7	765	13	ADS45896	Bacterial
C	40	161.4	3.7	906	11	ACH98155	Bacterial
C	41	159	3.7	792	11	ABD12618	Klebsiell
C	42	159	3.7	936	11	ABD12561	Pseudomon
C	43	158	3.6	742	13	ADT45738	Bacterial
C	44	156.8	3.6	876	11	ABD13397	Pseudomon
C	45	156.4	3.6	114955	2	AAV53491	Human ade

ALIGNMENTS

RESULT 1  
ABN86379  
ID ABN86379 standard; DNA; 4346 BP.  
XX  
AC ABN86379;

XX 21-OCT-2002 (first entry)

XX ScbA, ScbR and ScbB encoding genes containing DNA sequence.

XX Antibiotic; bacterium; scbA; afsA; scbR; arpA; barA; actinorhodine; Act;  
undecylprodigiosin; Red; gene; ds.

XX Streptomyces coelicolor.

XX CA2322241-Al.

XX 23-APR-2002.

XX 23-OCT-2000; 2000CA-02322241.

XX 23-OCT-2000; 2000CA-02322241.

XX (PLAN-) PLANT BIOSCIENCE LTD.

XX Takano E, Bibb M;

XX WPI; 2002-501089/54.

XX P-PSDB; AB880940, AB880941, AB880942.

XX Modifying antibiotic-producing Streptomycetes, to increase, or alter timing

XX of, antibiotic production, by deleting the scbA or scbR genes.

XX Claim 19, 20; Fig 14; 64pp; English.

XX The invention provides a method for modifying an antibiotic-producing  
strain of Streptomycetes to increase production of antibiotics or to alter  
the timing of antibiotic production. The modification is functional  
deletion of the scbA gene of S. coelicolor, or its homologues, but is not  
deletion of the afsA gene of S. griseus, or the modification is  
functional deletion of the scbR gene of S. griseus, or its homologues,  
but is not deletion of arpA of S. griseus nor barA of S. virginiae. The

CC method is particularly used for production of the antibiotics  
CC actinorhodine (Act) and undecylprodigiosin (Red). The present sequence  
CC represents the nucleic acid sequence of the region containing the S.  
CC cosicocolor ScbA, ScbR and ScbB proteins

Q Sequence 4346 BP; 650 A; 1561 C; 1519 G; 616 T; 0 U; 0 Other;  
Query Match 100.0%; Score 4346; DB 6; Length 4346;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4346; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy	1	GTCAGCAGCGGGTTCGGGTTTCGACGCCGACGCGGTACTCGTTTCCGGGCAACGGGACACGG	60
Db	1	GTCAGCAGCGCGTTCGGGTTTCGACGCCGACGCGGTACTCGTTTCCGGGCAACGGGACACGG	60
Qy	61	GGTCTGGGCTCGATGACACCGACCGCATCGAGGACGTCGGCTGGCGGCTCTCTGATGTGAGC	120
Db	61	GGTCTGGGCTCGATGACACCGACCGCATCGAGGACGTCGGCTGGCGGCTCTCTGATGTGAGC	120
Qy	121	GGCCCCGCGCGGCAACCATCGACATCGACGTCTCCACTGCGCCCCCGGAAAGTGAGC	180
Db	121	GGCCCCGCGCGGCAACCATCGACATCGACGTCTCCACTGCGCCCCCGGAAAGTGAGC	180
Qy	181	ACCGACCGCGGACGTGACGCCATGGGAGGGCCAGTCCGGGACGATCAACCTTGCT	240
Db	181	ACCGACCGCGGACGTGACGCCATGGGAGGGCCAGTCCGGGACGATCAACCTTGCT	240
Qy	241	TCGGCCGAAGGTTTCGGGTTTCGGCGCCCGACATGCGAACGGCCTGGGGGCGCGCC	300
Db	241	TCGGCCGAAGGTTTCGGGTTTCGGCGCCCGACATGCGAACGGCCTGGGGGCGCGCC	300
Qy	301	CGTCACTTCCGCAACGFTCGGACGACCTGCGGCTTGGCCCGCCCGCGGTGCGCTCGGC	360
Db	301	CGTCACTTCCGCAACGFTCGGACGACCTGCGGCTTGGCCCGCCCGCGGTGCGCTCGGC	360
Qy	361	GCTCTCAGGAAACGGTTCGGGACGCGGGCTCGGATCTTCGGAGCATCTGCTCGCGGC	420
Db	361	GCTCTCAGGAAACGGTTCGGGACGCGGGCTCGGATCTTCGGAGCATCTGCTCGCGGC	420
Qy	421	CGCCTCCAGCCCCAGTCTCTCGAAACGACGTGACCGCCCCAGCAGCCTCGGAGAGCGCGC	480
Db	421	CGCCTCCAGCCCCAGTCTCTCGAAACGACGTGACCGCCCCAGCAGCCTCGGAGAGCGCGC	480
Qy	481	CGTCTGCTCGGCGGTGATCGCTCGGGCCCGGACAGCGGGTACGCCGCTCGT	540
Db	481	CGTCTGCTCGGCGGTGATCGCTCGGGCCCGGACAGCGGGTACGCCGCTCGT	540
Qy	541	GCCGGGCCGGGTCAGTGCCACCGCCACATTCGCGCATGTCGCGCGATCGACGACGAC	600
Db	541	GCCGGGCCGGGTCAGTGCCACCGCCACATTCGCGCATGTCGCGCGATCGACGACGAC	600
Qy	601	AACCGGGACGTGCGGTACAGCGCGCGACACCGCGTCGACCGGATGGCGGGCGCCCA	660
Db	601	AACCGGGACGTGCGGTACAGCGCGCGACACCGCGTCGACCGGATGGCGGGCGCCCA	660
Qy	661	GGACAGGTGTTGGACATGAAGGTCCTGGCCCGCAGGAAGGTCAGTCTAGCCCGGACTC	720
Db	661	GGACAGGTGTTGGACATGAAGGTCCTGGCCCGCAGGAAGGTCAGTCTAGCCCGGACTC	720
Qy	721	CGGTACGGCCCGCTGTTCTCGCGTGGCGCGGTGATGAAGTCGTCGCGCCCGGTTTC	780
Db	721	CGGTACGGCCCGCTGTTCTCGCGTGGCGCGGTGATGAAGTCGTCGCGCCCGGTTTC	780
Qy	781	CCCCACCGCGAGCATGGACGTTTCCAGGTGTCGCGACGCGGCTTCGCGCGCCCGC	840
Db	781	CCCCACCGCGAGCATGGACGTTTCCAGGTGTCGCGACGCGGCTTCGCGCGCCCGC	840
Qy	841	CGCGAAACGCTCGTTCGGGCTCGGTGGCATGTTCTGTTGACGAGAAACGCGCCCGC	900
Db	841	CGCGAAACGCTCGTTCGGGCTCGGTGGCATGTTCTGTTGACGAGAAACGCGCCCGC	900
Qy	901	CCCGTTGAGGGCCCGGTCCAGGCCCGGGCGGTTCGGCGTACTCGCCCCGGGCGACCTCGAC	960



```
Db 4201 AAAAGCGCGCGCGGGTCCAGGGCTCTCCAGGAGCTTAGACCGCCAAACCGGCC 4260
Qy 4261 TGCGCCAGCGCGGACAGCGCCGCGCCAGCTCCGCGGTGTCGTCGCGCGCGCGAGGCG 4320
Db 4261 TGCGCCAGCGCGGACAGCGCCGCGCCAGCTCCGCGGTGTCGTCGCGCGCGCGAGGCG 4320
Qy 4321 AGCGACGGGTAGAGCGCGGGGTACC 4346
Db 4321 AGCGACGGGTAGAGCGCGGGGTACC 4346

RESULT 2
ADE14794
ID ADE14794 standard; DNA; 4346 BP.
XX
AC ADE14794;
DT 29-JAN-2004 (first entry)
XX
DE Streptomyces coelicolor partial genome sequence 2.
XX
KW antibiotic-producing strain; antibiotic production; scbA gene; ScbR gene;
KW actinorhodin; undecylprodigiosin; gene; ds.
XX
OS Streptomyces coelicolor.
XX
FH Key Location/Qualifiers
FT CDS complement(1199..2143)
FT FT /*tag= a
FT FT /product= "Streptomyces coelicolor SrbB protein"
FT FT /transl_except= (3149..3147, aa:Tyr)
FT FT 2261..2908
FT FT /*tag= b
FT FT /product= "Streptomyces coelicolor SrbR protein"
FT FT complement(3024..3812)
FT FT /*tag= c
FT FT /product= "Streptomyces coelicolor SrbA protein"
FT FT /transl_except= (3150..3145, aa:Met)
XX
PN US2003124644-A1.
XX
PD 03-JUL-2003.
XX
PF 23-OCT-2001; 2001US-00017471.
XX
PR 23-OCT-2000; 2000US-0242561P.
XX
PA (TAKA/) TAKANO E.
PA (BIBB/) BIBB M J.
XX
PI Takano E, Bibb MJ;
XX
PS WPI: 2003-810983/76.
DR P-PSDB; ADE14791, ADE14792, ADE14793.
XX
XX
XX Modifying an antibiotic-producing strain of Streptomyces coelicolor or
XX Streptomyces lividans to increase or alter the timing of antibiotic
XX production in the strain, comprises functionally deleting in the strain
XX the scbA or ScbR gene.
XX
XX Claim 19; Fig 14; 33pp; English.
XX
XX This invention relates to the novel modification of an antibiotic-
XX producing strain of Streptomyces coelicolor or Streptomyces lividans to
XX increase or to alter the timing of antibiotic production in the strain.
XX The method comprises functionally deleting in the strain the scbA or ScbR
XX gene. The method is useful in increasing and altering the timing of
XX antibiotic production (especially actinorhodin and undecylprodigiosin) in
XX Streptomyces species, particularly Streptomyces coelicolor or
XX Streptomyces lividans. The present sequence is that of a region of the
XX Streptomyces coelicolor genome, which encodes the ScbA, ScbB and ScbR
XX proteins and which is related to the invention.
```

```
XX
SQ Sequence 4346 BP; 650 A; 1561 C; 1519 G; 616 T; 0 U; 0 Other;
Query Match 100.0%; Score 4346; DB 10; Length 4346;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTCGACGACGCGCTCGGGTTTCGACGCGCGACGCGGTACTCGTTCCCGGCGCACCGGCGACCG 60
Db 1 GTCGACGACGCGGTTCGGGTTTCGACGCGCGACGCGGTACTCGTTCCCGGCGCACCGGCGACCG 60
Qy 61 GGTCTGCGCTCGATGACCGCATCGAGACGTCGCGTGGCGGTCTCTGATAGTGAGC 120
Db 61 GGTCTGCGCTCGATGACCGCATCGAGACGTCGCGTGGCGGTCTCTGATAGTGAGC 120
Qy 121 GGGCCCGCGCGCGACGACATCGAGCTCCATCTCCCATCTGCGCGCCCGCGGAAAGTGAGC 180
Db 121 GGGCCCGCGCGCGCGACGACATCGAGCTCCATCTCCCATCTGCGCGCCCGCGGAAAGTGAGC 180
Qy 181 ACCGACCGCGGACGTGACGCGCATGGGAGGGCCACGTCGCGGACGGATCACCCCTGGCT 240
Db 181 ACCGACCGCGGACGTGACGCGCATGGGAGGGCCACGTCGCGGAGCGATCACCCCTGGCT 240
Qy 241 TCGCGCGAAAGGCTTTCGCGTGGTCCGCGCCACAGATGCGGAAACGCGCTTGGCGGCGCGCC 300
Db 241 TCGCGCGAAAGGCTTTCGCGTGGTCCGCGCCACAGATGCGGAAACGCGCTTGGCGGCGCGCC 300
Qy 301 CGTCACTTCCGCGACGGTTCGGCAGCAGCTTGGCGCTTGGCCCGCGCGCTCGCGCTCGGC 360
Db 301 CGTCACTTCCGCGACGGTTCGGCAGCAGCTTGGCGCTTGGCCCGCGCGCTCGCGCTCGGC 360
Qy 361 GCTCTCAGGAAACGCTCGCGACGCGGCTCGGATCTTCGCGAGCATCTGCTCGCGCGC 420
Db 361 GCTCTCAGGAAACGCTCGCGACGCGGCTCGGATCTTCGCGAGCATCTGCTCGCGCGC 420
Qy 421 CGCTTCCAGCGCCAGCTCTCGAAACGCGAGTACCGCGCCCGCGACCTCGGAGAGCGCGC 480
Db 421 CGCTTCCAGCGCCAGCTCTCGAAACGCGAGTACCGCGCCCGCGACCTCGGAGAGCGCGC 480
Qy 481 CGTCTGCTGCTGGCGGTGATCGGCTCGGGCCCGGACGCGGTACGCGCGCTTCTCTGTCG 540
Db 481 CGTCTGCTGCTGGCGGTGATCGGCTCGGGCCCGGACGCGGTACGCGCGCTTCTCTGTCG 540
Qy 541 GCGGCGCGGTCAGTGCCCTGACCGCCTTCGCGATGTCGCGGATGTCGCGGATGTCGCGG 600
Db 541 GCGGCGCGGTCAGTGCCCTGACCGCCTTCGCGATGTCGCGGATGTCGCGGATGTCGCGG 600
Qy 601 AACCGGGGACGTGCGGTACAGCGCGCGACCAACGCGCTCGGACCGGATGGCGGGCGCCA 660
Db 601 AACCGGGGACGTGCGGTACAGCGCGCGACCAACGCGCTCGGACCGGATGGCGGGCGCCA 660
Qy 661 GGACAGCGTGTGGACATGAAGTCTCGGCGCGCGAGGATCCAGTCTAGCCCGGATC 720
Db 661 GGACAGCGTGTGGACATGAAGTCTCGGCGCGCGAGGATCCAGTCTAGCCCGGATC 720
Qy 721 GCGTACGCGCGCTGCTTCTCGCGCTCGCGGTGATGAAGTCTCGCGCGCGCGGTTTC 780
Db 721 GCGTACGCGCGCTGCTTCTCGCGCTCGCGGTGATGAAGTCTCGCGCGCGCGGTTTC 780
Qy 781 CCCCAACCGCGAGCATGGAAGCTTCAACAGGTGCGGACGCGCGGCTTCCGCGCGCGCGC 840
Db 781 CCCCAACCGCGAGCATGGAAGCTTCAACAGGTGCGGACGCGCGGCTTCCGCGCGCGCGC 840
Qy 841 CGCGAAACGCTGCTCGCTCGGCTCGGTGCGCATGTTCTGAGAGGAAACGCGCGCGCGAC 900
Db 841 CGCGAAACGCTGCTCGCTCGGCTCGGTGCGCATGTTCTGAGAGGAAACGCGCGCGCGAC 900
Qy 901 CCGGTTGAGGCGCGGTCACGCGCGCGGTGCGCGTACTCGCGCGCGGACGACCTCGAC 960
Db 901 CCGGTTGAGGCGCGGTCACGCGCGCGGTGCGCGTACTCGCGCGCGGACGACCTCGAC 960
Qy 961 GTTCCGCGCGGTGAGCGGTCAACCGTTCGCGCGCGCGGAGGACTTGTGACGGGACCGGT 1020
Db 961 GTTCCGCGCGGTGAGCGGTCAACCGTTCGCGCGCGCGGAGGACTTGTGACGGGACCGGT 1020
```







QY 1973 GGGCAGGACGGCGGTGACCGAGAACCGGTCAACCGCCCTTCGGTATCCAGCTGACCGGGAA 2032  
Db 1 GGGCAGGACGGCGGTGACCGAGAACCGGTCAACCGCCCTTCGGTATCCAGCTGACCGGGAA 60  
QY 2033 CGCGTCTCGACCCCTGGTCCGGTGGACCAAGCGCCATCGGAACCGGCAATGCGGTTTGTTC 2092  
Db 61 CGCGTCTCGACCCCTGGTCCGGTGGACCAAGCGCCATCGGAACCGGCAATGCGGTTTGTTC 120  
QY 2093 GATCGAGTTGGCATCGGACGAGAAATTGATCAAAACTACTGCTTCGGGCATGGGTCCCCC 2152  
Db 121 GATCGAGTTGGCATCGGACGAGAAATTGATCAAAACTACTGCTTCGGGCATGGGTCCCCC 180  
QY 2153 CCAGGAATCATGTGATGCGGAGCTGTTCTGTATGCGCGAACCTTAAGATACAGACTGAGC 2212  
Db 181 CCAGGAATCATGTGATGCGGAGCTGTTCTGTATGCGCGAACCTTAAGATACAGACTGAGC 240  
QY 2213 GGTGTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGAGGATGCGCCAGCAG 2272  
Db 241 GGTGTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGAGGATGCGCCAGCAG 300  
QY 2273 GACCGGGGATCCGACCGCGGACAGATCTCTGGACGCGCGCGCGAGCTCTTCGAGAAG 2332  
Db 301 GACCGGGGATCCGACCGCGGACAGATCTCTGGACGCGCGCGCGAGCTCTTCGAGAAG 360  
QY 2333 CAGGGCTACCAAGTGCACAGATCAACGAGATCTCAAGGT 2373  
Db 361 CAGGGCTACCAAGTGCACAGATCAACGAGATCTCAAGGT 401

## RESULT 4

ABN86378/c  
ID ABN86378 standard; DNA; 401 BP.  
AC ABN86378;  
XX  
DT 21-OCT-2002 (first entry)  
XX  
DE S. coelicolor ScbA and ScbR proteins encoding complementary DNA.  
XX  
KW Antibiotic; bacterium; scbA; afeA; scbR; arpA; barA; actinorhodine; Act;  
KW undecylprodigiosin; Red; gene; ds.  
XX

## Streptomyces coelicolor.

OS  
XX  
XX CA2322241-A1.  
XX  
XX 23-APR-2002.  
XX  
XX 23-OCT-2000; 2000CA-02322241.  
XX  
XX 23-OCT-2000; 2000CA-02322241.  
XX  
XX (PLAN-) PLANT BIOSCIENCE LTD.  
XX  
XX Takano E, Bibb M;  
XX WPI; 2002-501089/54.  
XX  
XX Modifying antibiotic-producing Streptomyces, to increase, or alter timing  
XX of, antibiotic production, by deleting the scbA or scbR genes.  
XX  
XX Disclosure; Fig 4B; 64pp; English.

XX  
CC The invention provides a method for modifying an antibiotic-producing  
CC strain of Streptomyces to increase production of antibiotics or to alter  
CC the timing of antibiotic production. The modification is functional  
CC deletion of the scbA gene of S. coelicolor, or its homologues, but is not  
CC deletion of the afeA gene of S. griseus, or the modification is  
CC functional deletion of the scbR gene of S. coelicolor, or its homologues,  
CC but is not deletion of arpA of S. griseus nor barA of S. virginiae. The  
CC method is particularly used for production of the antibiotics  
CC actinorhodine (Act) and undecylprodigiosin (Red). The present sequence  
CC represents a complementary fragment of the DNA encoding the S. coelicolor

CC ScbA and ScbR proteins  
XX  
SQ Sequence 401 BP; 72 A; 128 C; 114 G; 87 T; 0 U; 0 Other;  
Query Match 9.2%; Score 401; DB 6; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.3e-53;  
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1973 GGGCAGGACGGCGGTGACCGAGAACCGGTCAACCGCCCTTCGGTATCCAGCTGACCGGGAA 2032  
Db 401 GGGCAGGACGGCGGTGACCGAGAACCGGTCAACCGCCCTTCGGTATCCAGCTGACCGGGAA 342  
QY 2033 CGCGTCTCGACCCCTGGTCCGGTGGACCAAGCGCCATCGGAACCGGCAATGCGGTTTGTTC 2092  
Db 341 CGCGTCTCGACCCCTGGTCCGGTGGACCAAGCGCCATCGGAACCGGCAATGCGGTTTGTTC 282  
QY 2093 GATCGAGTTGGCATCGGACGAGAAATTGATCAAAACTACTGCTTCGGGCATGGGTCCCCC 2152  
Db 281 GATCGAGTTGGCATCGGACGAGAAATTGATCAAAACTACTGCTTCGGGCATGGGTCCCCC 222  
QY 2153 CCAGGAATCATGTGATGCGGAGCTGTTCTGTATGCGCGAACCTTAAGATACAGACTGAGC 2212  
Db 221 CCAGGAATCATGTGATGCGGAGCTGTTCTGTATGCGCGAACCTTAAGATACAGACTGAGC 162  
QY 2213 GGTGTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGAGGATGCGCCAGCAG 2272  
Db 161 GGTGTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGAGGATGCGCCAGCAG 102  
QY 2273 GACCGGGGATCCGACCGCGGACAGATCTCTGGACGCGCGCGCGAGCTCTTCGAGAAG 2332  
Db 101 GACCGGGGATCCGACCGCGGACAGATCTCTGGACGCGCGCGCGAGCTCTTCGAGAAG 42  
QY 2333 CAGGGCTACCAAGTGCACAGATCAACGAGATCTCAAGGT 2373  
Db 41 CAGGGCTACCAAGTGCACAGATCAACGAGATCTCAAGGT 1  
RESULT 5  
ADE14790  
ID ADE14790 standard; DNA; 401 BP.  
XX  
AC ADE14790;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Streptomyces coelicolor partial genome sequence 1.  
XX  
KW Streptomyces coelicolor partial genome sequence 1.  
KW antibiotic-producing strain; antibiotic production; scbA gene; ScbR gene;  
KW actinorhodin; undecylprodigiosin; gene; ds.  
XX  
OS Streptomyces coelicolor.  
XX  
XX  
XX Key Location/Qualifiers  
XX CDS complement(1..171)  
XX FT /tag= a  
XX FT /product= "Streptomyces coelicolor ScbA protein"  
XX FT /partial  
XX FT /note= "No stop codon"  
XX CDS 289..399  
XX FT /tag= b  
XX FT /product= "Streptomyces coelicolor ScbR protein"  
XX FT /partial  
XX FT /note= "No stop codon"  
XX  
XX US2003124644-A1.  
XX PN  
XX 03-JUL-2003.  
XX PD  
XX 23-OCT-2001; 2001US-00017471.  
XX PF  
XX 23-OCT-2000; 2000US-0242561P.  
XX PR  
XX (TAKA/) TAKANO E.  
XX PA



PA (BIBB/) BIBB M J.  
XX  
PI Takano E, Bibb MJ;  
XX  
DR WPI; 2003-810983/76.  
XX P-PSDB; ADE14791, ADE14792.  
XX  
PT Modifying an antibiotic-producing strain of Streptomyces coelicolor or  
PT Streptomyces lividans to increase or alter the timing of antibiotic  
PT production in the strain, comprises functionally deleting in the strain  
XX the scbA or ScbR gene.  
XX  
PS Example 4; Fig 4A; 33pp; English.  
XX  
CC This invention relates to the novel modification of an antibiotic-  
CC producing strain of Streptomyces coelicolor or Streptomyces lividans to  
CC increase or to alter the timing of antibiotic production in the strain.  
CC The method comprises functionally deleting in the strain the scbA or ScbR  
CC gene. The method is useful in increasing and altering the timing of  
CC antibiotic production (especially actinorhodin and undecylprodigiosin) in  
CC Streptomyces species, particularly Streptomyces coelicolor or  
CC Streptomyces lividans. The present sequence is that of a region of the  
CC Streptomyces coelicolor genome, which partially encodes the ScbA and ScbR  
CC proteins and which is related to the invention.  
XX  
SQ Sequence 401 BP; 87 A; 114 C; 128 G; 72 T; 0 U; 0 Other;  
  
Query Match 9.2%; Score 401; DB 10; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.3e-53;  
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1973 GGGCAGGACGGGTGACCGAGAACCGGTCAACCGCCCTTCGGTATCCAGCTGACCGGAA 2032  
Db 1 GGGCAGGACGGGTGACCGAGAACCGGTCAACCGCCCTTCGGTATCCAGCTGACCGGAA 60  
  
QY 2033 CGCGTCTCGACCCCTGGTCCGTGACAAAGCGCCATCGAACCGGCAATGCGTTTGTTC 2092  
Db 61 CGCGTCTCGACCCCTGGTCCGTGACAAAGCGCCATCGAACCGGCAATGCGTTTGTTC 120  
  
QY 2093 GATCGAGTTGGCATCGGACGAGCAATTCATCAAACTACTGCTTCGGGCATGGGTCCCC 2152  
Db 121 GATCGAGTTGGCATCGGACGAGCAATTCATCAAACTACTGCTTCGGGCATGGGTCCCC 180  
  
QY 2153 CCAGGAATCATGTGATCGCGAGCTGTTCTGTATCGCGCAACGTTAAGATACAGACTGAGC 2212  
Db 181 CCAGGAATCATGTGATCGCGAGCTGTTCTGTATCGCGCAACGTTAAGATACAGACTGAGC 240  
  
QY 2213 GGTGTTTTTTTCTATCTTCCGGGGGAGACATGAACAGGAGGCGAGGATGCCCAAGCAG 2272  
Db 241 GGTGTTTTTTTCTATCTTCCGGGGGAGACATGAACAGGAGGCGAGGATGCCCAAGCAG 300  
  
QY 2273 GACCGGGCGATCCGCACGCGGAGACGATCTCTGACGCGCGCGGCGAGGTCTTCGAGAAG 2332  
Db 301 GACCGGGCGATCCGCACGCGGAGACGATCTCTGACGCGCGCGGCGAGGTCTTCGAGAAG 360  
  
QY 2333 CAGGGCTACCAAGCTGCCACGATCAGGAGATCTTCAAGGT 2373  
Db 361 CAGGGCTACCAAGCTGCCACGATCAGGAGATCTTCAAGGT 401  
  
RESULT 6  
AAD47222  
ID AAD47222 standard; DNA; 19016 BP.  
XX  
AC AAD47222;  
XX  
DT 24-FEB-2003 (first entry)  
XX  
DE Streptococcus sp. H021 DNA fragment for rabelomycin biosynthetic pathway.  
KW Gene cluster; rabelomycin biosynthetic pathway; angucycline compound;  
KW drug screening; enzyme; gene; ds.  
XX

OS Streptococcus sp.  
XX  
FH Key Location/Qualifiers  
FT CDS complement(1..561)  
FT /\*tag= m  
FT /product= "Orfs"  
FT /note= "No stop codon"  
FT /partial  
FT complement(580..1335)  
FT /\*tag= n  
FT /product= "OrfR"  
FT complement(1391..2701)  
FT /\*tag= 1  
FT /product= "OrfQ"  
FT /transl\_except= (pos:2699..2701, aa:Met)  
FT complement(2712..3710)  
FT /\*tag= k  
FT /product= "OrfH"  
FT complement(3793..4848)  
FT /\*tag= s  
FT /product= "Orf2"  
FT complement(4854..5438)  
FT /\*tag= j  
FT /product= "OrfO"  
FT complement(5494..6687)  
FT /\*tag= o  
FT /product= "OrfY"  
FT complement(6780..8051)  
FT /\*tag= q  
FT /product= "OrfJ"  
FT complement(8178..8939)  
FT /\*tag= i  
FT /product= "OrfV"  
FT /transl\_except= (pos:8937..8939, aa:Met)  
FT complement(9014..10555)  
FT /\*tag= h  
FT /product= "OrfM"  
FT /transl\_except= (pos:10553..10555, aa:Met)  
FT complement(10557..11504)  
FT /\*tag= g  
FT /product= "OrfL"  
FT complement(11570..12352)  
FT /\*tag= d  
FT /product= "OrfD"  
FT /transl\_except= (pos:12350..12352, aa:Met)  
FT complement(12421..12684)  
FT /\*tag= c  
FT /product= "OrfC"  
FT complement(12681..13910)  
FT /\*tag= b  
FT /product= "OrfB"  
FT complement(13907..15142)  
FT /\*tag= a  
FT /product= "OrfA"  
FT /transl\_except= (pos:15140..15142, aa:Met)  
FT complement(15208..15537)  
FT /\*tag= f  
FT /product= "OrfP"  
FT complement(15586..17055)  
FT /\*tag= e  
FT /product= "OrfE"  
FT 17692..18492  
FT /\*tag= r  
FT /product= "OrfI"  
FT complement(18603..19016)  
FT /\*tag= p  
FT /product= "OrfR1"  
FT /note= "No start codon"  
FT /partial  
XX  
FN WO200274800-A1.  
XX  
XX 26-SEP-2002.  
PD





Db	12442	CCTGTCGTCA	12452	Matches	454;	Conservative	0;	Mismatches	291;	Indels	4;	Gaps	2;												
RESULT 8																									
ID	AA19362/c																								
XX	AA19362	standard; DNA; 7584 BP.																							
AC	AA19362;																								
DT	27-AUG-2003	(revised)																							
DT	19-MAY-1999	(first entry)																							
XX	Rhodococcus	corallina ohp operon.																							
XX	Rhodococcus	corallina; ohp operon; biosensor; mycolic acid bacteria;																							
KW	inducible	promoter; environmental pollutant; industry; medicine; ds.																							
XX	Rhodococcus	sp.																							
OS	W09900517-A2.																								
PN	07-JAN-1999.																								
PD	29-JUN-1998;	98WO-GB001893.																							
XX	27-JUN-1997;	97GB-00013666.																							
XX	(UYCA-)	UNIV CAMBRIDGE TECH SERVICES LTD.																							
PA	Archer JAC,	Summers DK, Roland HJ, Powell JAC;																							
PI	WPI; 1999-095760/08.																								
DR	P-PSDB; AAW99181,	AAW99182, AAW99183, AAW99184, AAW99185, AAW99186.																							
XX	Isolating DNA	encoding inducible promoter from mycolic acid bacteria -																							
PT	useful to produce	mycolic acid bacterial biosensors for particular																							
PT	analytes, such as	environmental pollutants, e.g. from industry or																							
PT	medicine.																								
XX	Example 7; Fig 4;	67pp; English.																							
PS	A method has been developed for identifying and/or isolating DNA from																								
CC	mycolic acid bacteria which encodes an inducible promoter induced in																								
CC	response to a particular analyte (and/or associated operon proteins). The																								
CC	method comprises: (a) culturing source of mycolic acid bacteria in																								
CC	selective media containing specific analyte and selective for																								
CC	oligotrophic bacteria; (b) identifying mycolic acid bacteria subsisting																								
CC	on medium; (c) extracting DNA from these bacteria; (d) incorporating																								
CC	into vector; (e) cloning vector into suitable host cell; and (f)																								
CC	screening host cells for inducible promoter and/or proteins to identify																								
CC	vectors encoding it. The method allows isolation of DNA encoding a																								
CC	promoter which is induced in response to a particular analyte (and/or																								
CC	associated operon proteins), which can be used to produce biosensors for																								
CC	the analyte. The DNA is incorporated into vectors which are used to																								
CC	transform host cells to produce the biosensors. It is especially useful																								
CC	when the analyte is an environmental pollutant (e.g. from industry or																								
CC	medicine), especially a hydrophobic organic compound such as components of																								
CC	fuels, pesticides; the biosensors may then be used to detect small																								
CC	concentrations of analytes in samples e.g. for pollution monitoring. The																								
CC	identified DNA can also be modified to produce modified inducible																								
CC	promoters and/or operons also useful in biosensor production. The method																								
CC	allows rapid isolation of promoters (and/or operon proteins) whilst																								
CC	minimising host restrictions (e.g. thick cell walls of mycolic acid																								
CC	bacteria which confer resistance to cell lysis) and requiring no																								
CC	knowledge of inducible enzyme chemistry involved (which may be unknown																								
CC	for specific analyte). The present sequence represents the R. corallina																								
CC	ohp operon, given in the present invention. (Updated on 27-AUG-2003 to																								
CC	correct OS field.)																								
XX	Sequence	7584 BP; 1329 A; 2735 C; 2286 G; 1234 T; 0 U; 0 Other;																							
SQ	Query Match	5.9%; Score 255.4; DB 2; Length 7584;																							
	Best Local Similarity	60.6%; Pred. No. 5.9e-31;																							
	Location/Qualifiers	296.1036																							
FH	Key	CDS																							
FT	CDS	296.1036																							

XX Rhodococcus sp. OHF operon.  
XX OHP operon; orthohydroxyphenylpropionic acid; regulator; inducible;  
KW gene expression system; plant transformation; ds.  
XX Rhodococcus sp; V49.  
XX Key Location/Qualifiers  
FH CDS 296.1036



CC derived from Streptomyces natalensis (ATCC 27448) which can be used for  
CC regulating the biosynthesis of natamycin. The novel polypeptides and  
CC their encoding polynucleotides can be used for mass-producing natamycin  
CC in microorganisms using an expression vector containing the natamycin  
CC gene and culturing the transformed microorganism in a medium. This  
CC sequence represents a DNA fragment containing a biosynthetic regulator of  
CC natamycin.  
XX  
SQ Sequence 4721 BP; 690 A; 1464 C; 1704 G; 797 T; 0 U; 66 Other;

Query Match		5.5%; Score 239; DB 14; Length 4721;
Best Local Similarity		50.3%; Pred. No. 2.2e-28;
Matches 876; Conservative 0; Mismatches 775; Indels 92; Gaps 8;		
QY	1213	CGGGGCGGACAGCGTGTGAGGAAGACGAGGCTGCGCTCTGATGCCCGGTGACCGCA 1272
DB	2678	CGGGGCGGTGAGCGTGGTGAACACCTTTGCGCGGTCTGTTGGCGGTGTCAGCGCA 2737
QY	1273	CGGTGTGACGCTGT 1332
DB	2738	CTGTGTGTGTGCGCCCTCACTTTCGCGCGCGCGG---CTGTGCTCCGATCCAGCACGGGCTGT 2794
QY	1333	CGAATCTCCGCTACCGGACGAAACCGGCTGCGCGGATCGAGGCGCAAGAGGCGCGGAC 1392
DB	2795	CGAATTCGACGTTCGAAAC---CGTGTGAGGCTGACGGCTAAAGGTCTTTTGTG 2851
QY	1393	CGGTACAGGCGACGCGCGCTGCGCTGCGCGCTCGAGACGACATGCGCGGTACGTGT 1452
DB	2852	CGGTACAGGCGCGCGCTGCTGCGCGCTTCGATCATGACGACCGGAAACGTGAT 2911
QY	1453	CGTTGGGCGCTGGAAGAGGGTGGGTGACGTGTGTGTCACCGCGAGTCGCGACGTGCTCT 1512
DB	2912	CGACGGGTGTGCGAACAGCGCGGTGTGCGGAAGTTCGCGCGCAGGAGCCACCGCGCG 2971
QY	1513	GCTGCGAATCTCGCGACAGACCACTCTCTGCGGGGTGACCGCGCGCGCGCGCGCG 1572
DB	2972	CGGTGTGCGGTGGGGTTCAGACCACTTCGACATCGCGCGGTGACGTCCACTGCGCGCG 3031
QY	1573	GCAAGGCGCGGTTCGCGGCGACCGATGCGGTGGGAGTTCGCGAAGT---GCGCGCGCATCC 1629
DB	3032	GGACGGGGCGCGGACGCGGCTGTGTGACGACCGCTCGCGGTGAGCGCGCGCGGCG 3091
QY	1630	GCGGTGAGACTTGAAGACTGTGAAGCGGTGCTGTGAGTTCCTGCGGTGCGACGAGCGGTC 1689
DB	3092	GCGGTGAGACCCCTGCGGGAATCCATTCGAAACCGGTGCGGTGCTGTGCGGACCTGCGCGC 3151
QY	1690	CGCGCGCGCGACCGCGCGCTGACCTGCTCCCTGTACGGGCTCCCGCGCGCGAATTC 1749
DB	3152	CGGCAAGTGGATCTCATCTGTGCGGACAGCGGTGATGGAACCGGCGACGCGACTGG 3211
QY	1750	GCTGGGAACAGGCGCACTTCACCTCCAGCTCCGCGACCTCGCGCGACACCGCGAGGTGT 1809
DB	3212	TGATTTGCGAGACAAAGTCAAGAACAGGCTGTGCTGGTTCCTTTCGCTCCAGGGCT 3271
QY	1810	CGAGGTGGAGGTGTAGTCAGCGTGGCCATCAAGAAAGTGTAGTCCCGCGACCGCGGT 1869
DB	3272	CCAGTTCGAAGGCGAAGTTCGAAGTTCGAGTTCGAGCAGCAGCGCATGCGTTCAGCGCGCGCG 3331
QY	1870	AGCGGGGTGAAGACAGGACATCGCGCTGACGCGAGGTCTCGGCGATCAGCAGCGGAT 1929
DB	3332	AACTCGGTGCGGAGCAGAGTCCGGTCTGCGCGACCATCTGCGGCGGCTTCATGCGCC 3391
QY	1930	CGGTGCGGTCCCGTGGACCGGTGCGAAGAACGGGTGTGCTGGGGCGAGGACGCGCGGTGA 1989
DB	3392	CAGTTCGGAACCGCGGCGCG---GGCGCGCGGTTCGCGCGGATGTCGCGGTGA 3448
QY	1990	CCGAGAACCGGTTCACCGCGCTTCGGTATCCAGTTCAGTTCAGCGGGAACGGCTCTTCGACCGCTGG 2049
DB	3449	CGCGGAATTCATGGCGCGCGTTCGAGTTCGAGTTCAGCAATACGAGTTCGCTGCTAAC 3508
QY	2050	TCCGTGTGAACAGGCGCATCGGAACCGGCAATGCGGTTCGATGAGTTCGATCGG 2109
DB	3509	GGAGGTGGAACGAATTCCTTGGGTACGCGCGACTGTGTCCGATGCGACGTGCGCGCGCG 3568

RESULT 11  
ADW38421/c

ID ADW38421 standard; DNA; 936 BP.

XX AC ADW38421;

XX DT 24-MAR-2005 (first entry)

XX DE DNA sequence related to Streptomyces nanchangensis.

XX KW Streptomyces nanchangensis; ds.

QY	2110	ACGCAATGATGATCAAACTACTGCTTCGGGCATGGTCCCCCGCAGGAATCATGTGATG 2169
DB	3569	GGAGCGCGCAGGTCTCTGTCGGGTGCGACAGGCTGTGCATACGGCCGAGCCCCAGGATG 3628
QY	2170	CGAGCTGTCTGTATGC-----GGGAACGTGA 2197
DB	3629	GTGTGACATAGTCCCTCCGTACCTCAAGCAACAGAGCTTCGGATGGCTCCCAATA 3688
QY	2198	AGATACAGACTGAGCGGTTTTTT-----TCT 2224
DB	3689	AGATACGTAATATCGGTTTGATTTTCAATGGGTGCGCACCGCGCTTCGGAACACTCG 3748
QY	2225	ATCTTCCCGGGGAGACATGAAACAGGAGGCA-----GGCATGCCAAGCAG 2272
DB	3749	TTTCAAGCAGGGGGGACCGCGCATGCGCGGTAAACATGAGAGATTGTCTATGGGAACAG 3808
QY	2273	GACCGGGCATCCGCAACCGCGCAGACGATCTCTGGACCGCGCGCGCAGGTCTTCGAGAAG 2332
DB	3809	GACCGGGCATACAGACCGCGCGCAGATCTCTGGATGCGCGGGCTTGTGTTCGACCAG 3868
QY	2333	CAGGGCTACAAAGTGCACGATCAACGAGATCTTCAAGTGGCCCGGGGTGACCAAGGGA 2392
DB	3869	CTCGGATACGAGTCCGCGACCATCACCGAGATCATCGCGAGCGCGCGGTGACGAAGGG 3928
QY	2393	GCCCTACTTCCATCTTCAGTCCAGGAAAGTGGCGCTGGCGCTTCGACGCCAG 2452
DB	3929	GCCCTGTACTTCCATCTTCAGTCCAGGAAAGTGGCGCGCGGTGCTCTCGAGGAGGCC 3988
QY	2453	GAACACCAACAGCGCTTCGCGAGCAACCCCTCGGCTGCAAGAACTCATTCGACATGGGC 2512
DB	3989	GTACACATGAGGGGTGCTACCGCAGGACTCCAAACTGACGAGGTGCGTGTGGG 4048
QY	2513	ATGTTGTTCTGTACCGCTTTCGACAGAACTGTGGCCCGCGCGCGGTGCGGCTCTCC 2572
DB	4049	ATGGTCTCGCCCAACCGTGTCCCGCGAGCGATGCTCAGTCCCGCATCCGCTGTCC 4108
QY	2573	ATGACACGAGCGCGACCGGTCTCGATCGCGAGGACCTTCGTCGCTGGCGACGAGACA 2632
DB	4109	ACCGACTTGAAGCCCGCTCTCTGTTCGCGACCT---CCTGGACGCCCTGGATCGCCTTC 4165
QY	2633	CTCCTGAAGCTGTGAACGAGGCAAGGAGAACGGTGTGCTGCGCCCATGTGGTCACC 2692
DB	4166	ATGGCAGACAGTGTGCGCTTCGCAAGGAGCAGGCGGAACTGTGCGCACACGTTCCG 4225
QY	2693	ACCGACTCGCGCGATCTCTACGTGGGACGTTTCGCGGGATACAGTGTGTCTCCAGACG 2752
DB	4226	GAGGAACCGCGTGTCTATCGTTCGCGCATGGAACCGGATGAGATCGTCACTCCGAC 4285
QY	2753	GTACGCGATACGAGGACCTCGAAGACCGGTACGCGCTGCTGCAAGGACATCTCTGCC 2812
DB	4286	CTCAAGGACGGGAGTCCCTTCAGCACCGGGTTTTCAGGCTCTACGACCATCTCTCTCCC 4345
QY	2813	GCCATCGCGTTCCTTCCTCGTGGCGCGCTCGATCTCTCCGAGGAGCGCGGACGACGC 2872
DB	4346	AGCATCGCGTTCGCGGTGTGTCGCGAGCTGAGTCACTACCGGACCGCGGTACCGG 4405
QY	2873	CTC 2875
DB	4406	ATC 4408



```
XX Streptomyces nanchangensis.
OS
XX Key Location/Qualifiers
FH 1..936
FT CDS
FT /*tag= a
XX
XX CN1472322-A.
FN
XX
XX PD
XX
XX PP 16-MAY-2003; 2003CN-00116989.
XX
XX PR 16-MAY-2003; 2003CN-00116989.
XX
XX PA (UYSH-) UNIV SHANGHAI COMMUNICATION.
XX
XX PI Deng Z, Sun Y, Zhou X;
XX
XX WPI; 2004-317409/30.
DR P-PSDB; ADM38422.
XX
XX Meilin mould C5 keto reductases gene, useful for reducing the C5-keto
XX group of the Meilingmycin glycoside.
XX
XX Claim 1; Page 2; 9pp; Chinese.
XX
XX The present sequence is a DNA sequence related to Streptomyces
XX nanchangensis.
XX
XX Sequence 936 BP; 153 A; 328 C; 308 G; 147 T; 0 U; 0 Other;
SQ
Query Match 5.5%; Score 238; DB 13; Length 936;
Best Local Similarity 59.7%; Pred. No. 3.4e-28;
Matches 472; Conservative 0; Mismatches 285; Indels 33; Gaps 3;
QY 3014 CGGTGGCGGTGACAGAGATGTTCCGCTGTGGCATCGACGTACTGCGCGGTGATCCACC 3073
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
925 GTGGTGTGGGGCCAGGGCCATGCTCCCGCGGAGCGCTGCACAGATGTTCCGGTGACCCAGC 866
QY 3074 GTGAGTCTGTGGAGGCCAGAAAGGCCACCACTGTCGGCGATGTCTGCGGTCTGCGGATGC 3133
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
865 GGGCTCTGTCGAGCGAGGAGGGCGACGACATCGCGGATGTGCGACGGCTGACCGAGTC 806
QY 3134 GATTGAACACGAGTTGGCGGCACTGTCGGCGCGCTGCGGGGTCTGCGCGCGCGCGTG 3193
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
805 GCGCGAGTGCCTGATGCGCGCCACGCGCTCTGCGCGCGCGCGGCATATCGTGACGCCACG 746
QY 3194 CGTTTCATGTCG ---TCTCCAGMAACCGCGCGCACCGCTGACCGTGATCCCGCGTT 3250
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
745 TGTGGTTTGTGTGTGTCGTTGATGCCCGCGCGATGTTGTGACGGTGATTCGCCCGT 686
QY 3251 CCCCCAGTTGCTGCGCCAGGGCGAGCGTGAGCGGTGTCCACCGCACCTTTGGTTCATCGCGT 3310
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
685 GTCCAGCTCATTTGGCCAGCTGCGGCGCATGATCTCGAGGGCGCCTTGTCTATGGCGT 626
QY 3311 ATCCGATGACCTCGGGGAACCGCGCGCGGTGCGCGGAGACGAGATGTTGATGATCCGCC 3370
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
625 A---GGCGACGACGCGCGCGAGCGATCCTGTTGACGCGGAGAGATGTTGACGATCCGTC 569
QY 3371 CGCGTTCGCGAGTGTGTTTCAGTTCGTGTCGACACGAGCGGTGCCCGAGTTGA 3430
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
568 CGCCGTGCGGAGGAGCGCGACACCGTTGGTGACGAAACCGGGGACGTAATTTGA 509
QY 3431 CGGCGACAGTCCGTCGAAGACTCTCTGGTGACTTCGTGATTCGTCGTCGCGAGCGCTGA 3490
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
508 CGCTGAAGATACGGTCGAACCTCTCTGATGACCTCGGAATGCGCCACATAGCGT 449
QY 3491 CGCCCGCGTTGTTACAGAGATGTGAACCTCGGCGGCGACTCCCGAACT----- 3538
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
448 CGCCCTGCGCGCGCGCGGTGCTCGCGGTGCTAGATGCGCGGTGTTGTTACACAGA 389
QY 3539 -----CGCCCATCCCGCGGTGCAACGCGCGTGAAGCGCGCGCGGTGCAC 3583
```

```
DB 388 TGTTCAGCGCGCGCGCGAGCCCGTCTCGAGCTCGTGAAGCGTGTTCGAGGCGGT 329
QY 3584 CCACGAGCCGAGTTTCGCGCCCGGATGCGCAACGCGCTGTTCGCCCGTGTCTCGGATGTCT 3643
DB 328 CGTGGCGCGAAGGTTCGCGCCCTGATGCGCAATGCTGCGCGCGCGCTGTCTCGATCGACG 269
QY 3644 CGACGCTCTCTCGCGCGCGCGCTCGTCTGCTGCGGTAGTGGACTGCGACGAGCGCCCGT 3703
DB 268 CGACCGTTTCTTCGCGCGTCTTTCACCGCCGTAGTCACGCGCCACCAAGAACACCTT 209
QY 3704 CCGCGCGCAGCGCGAGCGCGATACCGGCTTCGATGCGCGCGCTTCCCGCGTCCACGAGG 3763
DB 208 CGGCGCGAGCGCTTCGGGCGATGCGCGCGCGATGCGCGCGCGCGCGCGCGCTCACCAGTG 149
QY 3764 CGGTCTTGCC 3773
DB 148 CGGTTTTTCC 139
RESULT 12
AAL44719
ID AAL44719 standard; DNA; 699 BP.
XX
AC AAL44719;
XX
XX 03-MAY-2002 (first entry)
XX
XX S virginiae butanolide binding protein coding sequence.
XX
XX Butanolide binding protein; gene expression induction; operator;
XX transgenic plant; antibiotic production; gene; ds.
XX
XX Streptomyces virginiae.
XX
XX Key Location/Qualifiers
FH CDS 1..699
FT /*tag= a
FT /product= "butanolide binding protein"
XX
XX WO200196581-A1.
XX
XX 20-DEC-2001.
XX
XX 15-JUN-2001; 2001WO-JP005096.
XX
XX 15-JUN-2000; 2000JP-00180466.
XX
XX (KANF ) KANEKA CORP.
XX
XX Shinmyo A, Kato K, Yamada Y, Nihira T, Shindo T;
XX
XX WPI; 2002-098073/13.
XX
XX P-PSDB; AAM48990.
XX
XX Inducing expression of gene under regulation by operator at actinomycetes
XX self-regulator provision site, useful in producing transformant tobacco
XX for production of antibiotics e.g. virginiamycin.
XX
XX Example 1; Page 46-48; 57pp; Japanese.
XX
XX The present invention relates to a method of inducing the expression of a
XX gene under the regulation of an operator at an actinomycetes self-
XX regulator provision site comprising imparting the characters of a
XX repressor and an operator constituting a gene expression-inducing system
XX with use of the actinomycetes self-regulator as inducer to a plant by
XX gene transfer thereby providing the self-regulator to the transformed
XX plant..The method is useful in producing transformant tobacco for
XX production of antibiotics e.g. virginiamycin. The present sequence is the
XX Streptomyces virginiae butanolide binding protein coding sequence which
XX was used in the exemplification of the invention
XX
XX Sequence 699 BP; 116 A; 233 C; 253 G; 97 T; 0 U; 0 Other;
```



```
Query Match      5.3%; Score 229; DB 6; Length 699;
Best Local Similarity 60.8%; Pred. No. 8.7e-27;
Matches 373; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

QY 2270 CAGGACCGGGCGATCCGACCGGGCAGACGATCTTGAGACCGCGCGCAGGTCTTCGAG 2329
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34 CAGGAAACGGCGGTCCGACCGGGCAGGCGATCGTGGCGGCGACGCCCTCGGTCTTCGAC 93
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2330 AAGCAGGGCTACCAAGCTCCAGATCCTCAGGATCCTCAAGGTGCGCGGGTGACCAAG 2389
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94 GAGTACGGGTTCGAGCGCGCCACAGTGGCAGAGATCTCTCGGGGCTCGGTACCAAG 153
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2390 GGAGCCCTCTACTTCCACTTCCAGTCCAGAGGAAGAACTCGCGCTGGCGGTCTTCGAGCGC 2449
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154 GCGCGATGTACTTCCACTTCCAGTCCAGAGGAAGAGTGGCGCGCGGTCTGGCGGAG 213
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2450 CAGGAACACACACAGCGCTTCGAGAGCAACCCCTCGGCTCGAGAACTCATGACATG 2509
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 214 CAGACCTTCGACGTGGCGGTTCGCGGAATCCGGCTCCAGGGCGCAGGAATGGTAGACCTC 273
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2510 GGCATGTTGTTCTTCCAGCTTTCGACAGACGTGTCGCGCGCGCGCGGTGCGCCTC 2569
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 274 ACCATGCTGTGCTCCCGCAGCGATCTGACGATCCGATCTCGCGCGCGCGCGGTCTC 333
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2570 TCCATGACACAGCAGCGCAGCGTCTCGATCGCGAGGACCCCTTCGCTGCTGGCACGAG 2629
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 334 GCACTGGACCGGGGCGGTGGACTTCTCGACGCCAACCCGTTCGCGAGTGGGCGGAC 393
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2630 ACATCTCTGAAGTGTGAACACAGGCGCAGGAGAGCGTGTGCTGCTGCCCATGTGGTC 2689
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 394 ATCTGCGCCCGCTCTCGCGAGGACAGGAACGGGGGAGGTGCTTCGCGCAGTGAAC 453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2690 ACCACGACTCGGCCGATCTTACGTGGGCACTTCGCGGGGATACAGTCTGTGCCAG 2749
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 454 CCGAAAGAACCGGCACTTTCATGCTGGTGTCTTACCGGGCTCAGGGGCTCTCCGG 513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2750 ACGGTTCAGGACTTACAGGACCTTCGAACACCGCTACGCGTCTGTCAGAGGACATCTCTG 2809
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 514 GTCACCTTCGACCGCCAGGACCTTCGCGCACCGGATCTCGGTGATGTGAACACAGTGTG 573
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2810 CCGGCATCGCGTTCCTTCGCTGTCGGCGCGCTGATCTTCGAGGAGCGCGAGCA 2869
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 574 CCAGCATGTCGCGCGTCTCATGCTGACCTGTGATCGAAACCGCGAGGAGCGGATCGGG 633
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2870 CGCTCGCGCGC 2882
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 634 AAGTTCGCGCGG 646
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 13
AC164643/c
ID AC164643 standard; DNA; 14638 BP.
XX
XX AC164643;
XX AC
XX AC
DT 02-JUN-2005 (first entry)
XX
DE M. xanthus DNA fragment, seq id 1106.
XX
XX Transgenic plant; DNA replication; gene regulation; gene expression; ds.
XX
XX Myxococcus xanthus.
XX
XX US6833447-B1.
XX
XX 21-DEC-2004.
XX
XX 10-JUL-2001; 2001US-00902540.
XX
XX 10-JUL-2000; 2000US-0217883P.
XX
XX (MONS ) MONSANTO TECHNOLOGY LLC.
XX
```

```
XX Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX WPI; 2005-028716/03.
XX
XX New substantially purified Myxococcus xanthus nucleic acid molecule
XX encoding a nitrite reductase, useful for determining gene expression,
XX identifying mutations in a gene of interest, and for constructing
XX mutations in a gene of interest.
XX
XX Example 1; SEQ ID NO 1106; 25pp; English.
XX
XX The invention relates to a substantially purified nucleic acid molecule
XX encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
XX recombinant DNA construct for expression of a nitrite reductase gene in a
XX plant cell, and a plant cell comprising the recombinant DNA construct.
XX The nucleic acid is useful for determining gene expression, identifying
XX mutations in a gene of interest, and for constructing mutations in a gene
XX of interest. Sequences given in records for SEQ IDs 1-1849 represent a
XX set of 1849 contig and singleton sequences comprising coding sequences,
XX DNA replication elements, promoters and other regulatory elements from
XX the genome of the bacterium Myxococcus xanthus. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from USPTO
XX
XX Sequence 14638 BP; 2293 A; 4640 C; 5266 G; 2439 T; 0 U; 0 Other;
XX
XX Query Match      5.1%; Score 219.8; DB 14; Length 14638;
XX Best Local Similarity 56.5%; Pred. No. 2.1e-25;
XX Matches 429; Conservative 0; Mismatches 327; Indels 3; Gaps 1;

QY 3009 CCGAAGCTGGCGGGTTCAGAGAATCTTCCGCTGTGGATCGACGTACTGGCGGTGAT 3068
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12645 CTGGAGCGCGCGCTTCAGAGCATCATCCGCGCGAGGGCTCGATCGTGGCGGTCTAC 12586
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3069 CCACCTGTGATGTCGGAGGCCAGAAAGCCACAGTGGCGGATCTGTCGGGTCTGCC 3128
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12585 CCACGCGCTCTCGGGGACAGCAGCATGGCAGCCGATCCCCGATGTCGCGGCTGAC 12526
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3129 GATGCGGTGTAACACAGCGAGTTCGCGGCCAGTTCGCGCGCGCTCGGGGGTCTGCGCGCG 3188
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12525 CACAGCGCGAGGCAACCTGGCGCGCCAGCGCTTGTTCAGCTCCGGGTGTGACGGAC 12466
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3189 CGGTGCGTTC---ATGTCGCTTCCAGAAACCGCGCGCCACCGGCTGTGACCGGTATCCC 3245
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12465 GACGCGCGCCCGAAGTTCGCGTTCGCGCTCGCGCGCGCGGAGCACCTTGACGGAATCCT 12406
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3246 CGGTTCGCCCGAGTTCGCGCGAGGCGAGCGTGTCCACCGACCTTGGTTCAT 3305
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12405 CCGGGCCCCAGCTCTTCGCGAGGTACCGGGTGAGGCGCTCCACGCCCGCTTCATGGT 12346
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3306 CCGTATCCGATGGAATTCGCGGGAACCGCGCGCGCGGTTCGCGGACAGAGATGTTGATGAT 3365
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12345 CCGGTAGGCGGGTTCGCGAGGAGGTGAAGCGCGCGAGCGCGGAGGAGATGTTCCAGAT 12286
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3366 CCGCGCGCGCTCGCGCAGTGTTCAGTTCGCTGTCGACAGAAACAGCGGTGCCCGGAC 3425
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12285 GCGCGCCCCATTCGCGCAGCACCGCAACAGCGCTCGGTGAGGAAGAACGCTCCCTTTGAG 12226
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3426 GTTGACGGCGACCGTTCGTCGAGACCTTCCTCGGTGACTTCGCTGATCCGTCGCGAGCC 3485
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12225 GTTGACGTTTCATCAGCGCGTTCGAACTGGGCTCGGTCTCGGTGAGGAGTGTGAT 12166
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3486 GCTGACGCGCGCTGTTTCCAGAGGATGTCGAATCTCGGGCGGCACTCCCGAACTCGGCCAT 3545
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12165 GCCATGCGCGGTGTTTCCAGAGGATCCAGCGCTCGCGCGCGGAGTGTGACGGCCAG 12106
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3546 CCGCGGTTCGAACCGCGGTGTAGAGCGCGCGGTCAACCCACAGCGCGGATGTCGCGCG 3605
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12105 CTCGTCGCGCACCGCTTCCAGAAAGCGCGCGGAGCGCGGTGTCCCGCCACGTCGAGCGG 12046
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3606 GATGGCCAAAGCTGTTCGCGCGCTGCTCCGATGGTTCGACGCTCTCTCGCGCGCGCG 3665
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 12045 GAGCGCCACCGCTTGGCCGCTGCTCTCAATCTGCTTACAGACCGCGCGCTTCGTC 11986  
Qy 3666 CTCGTTGTCGCGTAGTGGACTGCCACGAGCCCGCTCCGGCGCCAGCGAGGGCGAT 3725  
Db 11985 CCGCGCGGTGCGGTACGTACGATGATCGCGTGGCCCGCGCCAGCTTGAGCGCCAT 11926  
Qy 3726 ACCGGGTGCGATGCGCCCGCTTCCCGCGGTACACGAGGC 3764  
Db 11925 GTTCCGCCCATGTCACGGCTTCCCGCGGTGATGAGC 11887

RESULT 14  
ACL65960/c  
ID ACL65960 standard; DNA; 744 BP.  
XX  
AC ACL65960;  
XX  
DT 02-JUN-2005 (first entry)  
DE M. xanthus gene sequence, seq id 2423.  
XX  
KW Transgenic plant; DNA replication; gene regulation; gene expression;  
KW gene; ds.  
XX  
OS Myxococcus xanthus.  
XX  
FN US6833447-B1.  
XX  
PD 21-DEC-2004.  
XX  
PF 10-JUL-2001; 2001US-00902540.  
XX  
PR 10-JUL-2000; 2000US-0217883P.  
XX  
PA (MONS ) MONSANTO TECHNOLOGY LLC.  
XX  
PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;  
XX WPI; 2005-028716/03.  
XX  
PT New substantially purified Myxococcus xanthus nucleic acid molecule  
PT encoding a nitrite reductase, useful for determining gene expression,  
PT identifying mutations in a gene of interest, and for constructing  
PT mutations in a gene of interest.  
XX  
PS Example 2; SEQ ID NO 2423; 25pp; English.

XX  
CC The invention relates to a substantially purified nucleic acid molecule  
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a  
CC recombinant DNA construct for expression of a nitrite reductase gene in a  
CC plant cell, and a plant cell comprising the recombinant DNA construct.  
CC The nucleic acid is useful for determining gene expression, identifying  
CC mutations in a gene of interest, and for constructing mutations in a gene  
CC of interest. Sequences given in records for SEQ IDs 1850-9691 represent a  
CC set of about 7842 genes or partial genes from the genome of the bacterium  
CC Myxococcus xanthus. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from USPTO  
XX  
SQ Sequence 744 BP; 113 A; 235 C; 287 G; 109 T; 0 U; 0 Other;

Query Match 5.0%; Score 216; DB 14; Length 744;  
Best Local Similarity 56.6%; Pred.No. 9.3e-25;  
Matches 421; Conservative 0; Mismatches 320; Indels 3; Gaps 1;  
Qy 3024 TCAGAGATCGTTCGCGCTGTCGCGATCGACGTACTGCGCGGTGATCCACCGTGAATCGTC 3083  
Db 744 TCAGAGCATATCCCGCCGAGGCTCGATGCGTGGCCCGCTCACCCAGCGCTCTCCGG 685  
Qy 3084 GGAGCCGAGAAAGGCCACACGTCCGGCGATGTCGTGGGTCTGCGCGATGCGTTGAACAC 3143  
Db 684 CGACAGCAGCATGGCGACCGCATCCCGCTCGTGGGTCTGACCCACACGCGCGCGC 625

Qy 3144 GGAGTTGGCGCCAGTCGCGCGCGCGCTTCGCGGGTCTGCGCGCGCGCTGCGTTTC---AT 3200  
Db 624 AACCTGGCGCGCCAGCGCTTGTTCAGCTCCGGGTTGTACGGACGACGCGGCCCGCGAA 565  
Qy 3201 GTCCGTCTTCCACGAAACCCCGGGCACCGCGTTGACGTGATCCCGTTCCCGCCAGTTG 3260  
Db 564 GTCCGTTCGCGTTCGCGCGCGCGCGAGCACGTTCAGCGAAATCTCTCCGGCGCGCCAGCTC 505  
Qy 3261 CCTCGCCAGGGCGAGCGGTGTCACCGCACCCCTTGGTTCATCGGTATCCGATCGA 3320  
Db 504 CTTTCGAGGTTACCGGGTGAGCGCTTCCACGCCCGCTTCATGGTCGGCGGCGGTG 445  
Qy 3321 CTCGGGGAACCGCGCGCGGTTCGCGCGAGAGAGATGTTGATGATCCCGCCCGCTCGCG 3380  
Db 444 GCCAGGGAAGTGAAGCGCGAGCGCCGAGAGATGTTTCAGGATGCGCCCCCATCCGC 385  
Qy 3381 CAGTCGTTTCAGTCGTCTGGACACGAGAACAGCGGTGCCCGAGCTTGACGGCGACGAG 3440  
Db 384 CAGCACCGGCAACAGCGCTCGTGGAGGAGAACTCCCTTTCAGGTGAGCTTCATCAG 325  
Qy 3441 TCGGTTCGAGAACCTCTCGGTGACTTCCGTGATCCGTCCCGAGCGCTGACGCCCGCTT 3500  
Db 324 CGCGTCGAACTGGGGCTTCGGTCTCCGCGAAGTGAATTCGATGATGCGCCGCTT 265  
Qy 3501 GTTCACAGGATGTCGAACTCGGGCGCGCACTCCGAACTCGCCCATCCCGGGGTTCGAAACGC 3560  
Db 264 GTTCACAGGAACTCCAGCGCTCGCGCCGAGTGACGGCCAGCTCCGTCGCGACCGC 205  
Qy 3561 CGGTAGAGCGCGCGCGCTCACCCACAGCCGAGTTTCGGCCCGGATGAGCGCAACGCTG 3620  
Db 204 TTCACAGAGGCGCGAAAGCCCGCTGTCGCCACGTCGAGCGGGAGCGCACGCCCTT 145  
Qy 3621 TCGCGCGCTGCTCCGGATGCTTCGAGGTTCTTCGCGCGCGCGCTCGCTGCTGCTGCTA 3680  
Db 144 GCCGCTGCTGCTTCAATCTGCTTCAACGACCGCGCGCTTCGTCGCGCGCGGTGCGGTA 85  
Qy 3681 GTGACTGCCACGAGCGCGCTCCGCGCGCAGCGCGAGGGCGATACCGCGTCCGATGCC 3740  
Db 84 CGTCAGGATGATGCGCGTCCCGGGCGCGCGCTTGAGCGCCATGTTCCGCCCCAGTCC 25  
Qy 3741 CCGGCTTCCCGCGTACACAGGC 3764  
Db 24 ACGGCTTCCCGCGTGAAGAGC 1

RESULT 15  
ADT44601/c  
ID ADT44601 standard; cDNA; 741 BP.  
XX  
AC ADT44601;  
XX  
DT 02-DEC-2004 (first entry)  
XX  
DE Bacterial polynucleotide #19352.  
XX  
KW Recombinant DNA construct; transformed plant; improved plant property;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
KW pathogen tolerance; pest tolerance; plant disease resistance;  
KW cell cycle pathway modification; plant growth regulator;  
KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
KW bacterial polynucleotide; gene; ss.  
XX  
OS Bacteria.  
XX  
FN US2003233675-A1.  
XX  
PD 18-DEC-2003.  
XX  
PF 20-FEB-2003; 2003US-00369493.  
XX  
PR 21-FEB-2002; 2002US-0360039P.  
XX

PA	(CAOY/) CAO Y.	
PA	(HINK/) HINKLE G J.	
PA	(SLAT/) SLATER S C.	
PA	(CHEN/) CHEN X.	
PA	(GOLD/) GOLDMAN B S.	
XX		
PI	Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;	
XX		
DR	WPI; 2004-061375/06.	
XX		
PT	New recombinant DNA construct comprising a promoter positioned to provide	
PT	for expression of a polynucleotide encoding a polypeptide from a	
PT	microbial source, useful for producing plants with improved properties.	
PS	Claim 1; SEQ ID NO 43039; 122pp; English.	
XX		
PS	The invention relates to a recombinant DNA construct comprising a	
CC	promoter functional in a plant cell, where the promoter is positioned to	
CC	provide for expression of a polynucleotide encoding a polypeptide from a	
CC	microbial source. The invention also relates to a transformed plant	
CC	comprising the recombinant DNA construct and a method of producing a	
CC	transformed plant having an improved property. The plant is a crop plant	
CC	such as maize or soybean. The method of producing a transformed plant	
CC	having an improved property comprises transforming a plant with the	
CC	recombinant DNA construct and growing the transformed plant, where the	

Search completed: January 18, 2006, 10:27:13  
Job time : 2379 secs

Blank Page (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2006, 09:45:41 ; Search time 21074 Seconds  
(without alignments)  
11722.579 Million cell updates/sec

Title: US-10-017-471B-19  
Perfect score: 4346  
Sequence: 1 gtgcagcagcggtcggtt.....gcgtagagcgcggtacc 4346

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_in.\*  
3: gb\_env.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pr.\*  
9: gb\_ro.\*  
10: gb\_sts.\*  
11: gb\_sy.\*  
12: gb\_un.\*  
13: gb\_vi.\*  
14: gb\_htg.\*  
15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4346	100.0	6727	1	SCO007731
2	4299.8	98.9	290850	1	SCO939127
3	901.2	20.7	12070	1	AB011413
4	563	13.0	210614	1	AB088224
5	527.4	12.1	1203	1	STWAFSAA
6	441.8	10.2	6252	1	SVU24659
7	377.2	8.7	110000	1	BA000030.43
8	347.2	8.0	5495	1	AY228176S2
9	335	7.7	290850	1	SCO939127
10	322.8	7.4	17455	1	AY956334
11	322.6	7.4	16643	1	AY034378
12	316	7.3	39428	1	AF322256
13	314.4	7.2	34869	1	AF324838
14	305.2	7.0	343243	1	BX640414
15	303.8	7.0	348997	1	BX640427
16	303.6	7.0	348706	1	BX640445
17	302.6	7.0	8018	1	AF164960
18	289.8	6.7	92294	1	AY117439

19	280	6.4	2234	1	AB001683	AB001683	Streptomy
C	20	279.4	6.4	34644	1	AF080235	Streptomy
C	21	277.6	6.4	1215	1	AB001608	Streptomy
C	22	276.8	6.4	300100	1	SCO939123	Streptomy
C	23	274.2	6.3	88400	6	CQ924550	Sequence
C	24	255.4	5.9	7584	6	AB1684	Sequence 1
C	25	255.4	5.9	7584	6	AB1685	Sequence 2
C	26	255.4	5.9	7584	6	ARG33953	Sequence
C	27	255.4	5.9	7584	6	ARG33954	Sequence
C	28	255.4	5.9	7599	1	AF274045	Rhodococ
C	29	255.4	5.9	7600	6	AR437655	Sequence
C	30	255.4	5.9	7600	6	AX026699	Sequence
C	31	253.6	5.8	110000	1	BA000030.24	Continuation (25 o
C	32	242.2	5.6	5998	1	AB032524	Streptomy
C	33	242.2	5.6	110000	1	BA000030.11	Continuation (12 o
C	34	239.4	5.5	302007	1	SCO939132	Streptomy
C	35	238	5.5	936	1	AY262284	Streptomy
C	36	229	5.3	699	6	BD105336	Method fo
C	37	229	5.3	1755	1	STMBARA	Streptomyce
C	38	225.4	5.2	32870	1	AF007101	Streptomy
C	39	219.8	5.1	14638	6	AR619711	Sequence
C	40	216	5.0	744	6	ARG21022	Sequence
C	41	209.6	4.8	2234	1	AB001683	Streptomy
C	42	209	4.8	2232	1	AY640376	Streptomy
C	43	209	4.8	2637	1	AF156161	Streptomy
C	44	208.4	4.8	10480	1	AE008375	Agrobacte
C	45	208.4	4.8	10575	1	AE009238	Agrobacte

ALIGNMENTS

RESULT 1  
SCO007731  
LOCUS Streptomyces coelicolor scbR gene, 6727 bp DNA linear BCT 15-APR-2005  
DEFINITION Streptomyces coelicolor scbR gene, scbA gene, ORFs A,B,X & Z.  
ACCESSION AJ007731  
VERSION AJ007731.1 GI:3425857  
KEYWORDS gamma-butyrolactone binding protein; scbA gene; scbR gene.  
SOURCE Streptomyces coelicolor A3(2)  
ORGANISM Streptomyces coelicolor A3(2)  
REFERENCE 1  
AUTHORS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetes.  
TITILE Takano,E., Chakraborty,R., Nihira,T., Yamada,X. and Bibb,M.  
JOURNAL Characterisation of scbR, and scbA genes involved in gamma-butyrolactone binding and synthesis in Streptomyces coelicolor  
AUTHORS Unpublished  
TITLE 2 (bases 1 to 6727)  
JOURNAL Direct Submission  
AUTHORS Takano,E.  
TITLE Submitted (11-AUG-1998) Takano E., Dept. Genetics, John Innes Institute, Colney lane, Norwich, NR4 7UH, UK  
JOURNAL Location/Qualifiers  
FEATURES  
source  
1..6727  
/organism="Streptomyces coelicolor A3(2)"  
/mol\_type="genomic DNA"  
/strain="M145"  
/db\_xref="taxon:100226"  
37..969  
/gene="orfB"  
37..969  
/gene="orfB"  
/codon\_start=1  
/transl\_table=11  
/product="histidine kinase"  
/protein\_id="CAA07625.1"  
/db\_xref="GI:3425858"  
/db\_xref="GOA:O86849"  
/db\_xref="InterPro:IPR000209"  
/db\_xref="InterPro:IPR003594"  
/db\_xref="InterPro:IPR011712"

```
/db_xref="UniProt/TREMBL:O86849"
/translation="MGTPSPENRRGGLLIPCSLDPRVSRRTVERHFCPTPARHD
LAGSHRGRTFTIQLDGLSRFRGRARQTTGTAHPGLPAGRDGPAAGPVRTDRGRDD
HWAARLRAREDELGTALSTAAHHIELHAAETGGTHLDAARHSLREMAVARRL
TGELQAOVTPLPLAQAVEFAATTPGCTEVRITRTGDRLLSDVCRRELFVAREAL
HNAFHADRVTVTLTRTRWAHAGIVDDGVGFDADAVLVFGHRAPLGRSMTDRIED
VGMRLIVSGPAGGTHIDVHLPLPRKVVSTAPRI"
complement(1000. .1854)
/gene="orfA"
complement(1000. .1854)
/gene="orfA"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAA07626.1"
/db_xref="GI:3425859"
/db_xref="UniProt/TREMBL:O86850"
/translation="MWLVGTGVGVQVAVHLLARTGPVRVLARRPERVTVTGNVEV
CAGYADRPGLDRALNGVRAAFVLTNSATEPDDRFPAARAEAGVRLVKLSMLVGE
PGADDFITRQRENERAVRESGLDWTFLRAKTFMSNTLSWAPAIRSDGVVRALYGTSP
VACVDPRDIAEVAVALTRPGHGRAYALSQPEAITAQQTARLSVLGSLRPEELG
LEAAREOMLRYPVPVADAFLESAERQAGAKAQVVTREVTGRPARPFIWAADHA
BAFCRSQG"
complement(1970. .2914)
/gene="scbA"
complement(1970. .2914)
/gene="scbA"
/function="involved in gamma-butyrolactone synthesis"
/codon_start=1
/transl_table=11
/product="ScbA protein"
/protein_id="CAA07627.1"
/db_xref="GI:3425860"
/db_xref="UniProt/TREMBL:O86851"
/db_xref="UniProt/TREMBL:O86852"
/translation="WPEAVLINSASDANSIEQTALPVPMLVHRTRVQDAPVPSWIP
KGRDFTAVLPHDHPFPAVHGDRDPLLIATLROAQMVLVHAGVGVPGVHFLM
ATLDYTHLDHGLVGEVALEVEVACSQLPRGGQVQGVQVQVDMWARRAGLAAIGTA
TTRFTSPQVYRMGRGDFATPTASVPGTAPVPAARGRTRDEDVVLSSQQDTWRLRV
DTSHTPLQRPNDHVPQMLLEAARQAACLVTGPAPFVPSIGSTRFVRYAEFDSPCWI
QATVRPGPAAGLTTRVTGHQDGLVFLTLTSLGPAFSG"
3032. .3679
/gene="scbR"
3032. .3679
/gene="scbR"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="gamma-butyrolactone binding protein"
/protein_id="CAA07628.1"
/db_xref="GI:3425861"
/db_xref="GOA:O86852"
/db_xref="UniProt/TREMBL:O86851"
/db_xref="UniProt/TREMBL:O86852"
/translation="WAKQDRALTRQTIIDRAAAQVFEKQVQAATITELKXVAGTKG
ALYFFHQSKBALGVDAQPEPPQVPEQPLRLQELIDMGLMFLHRLTNVVARAGVR
LSMDQAHGLDRGRGFRFHWETLLKLNAKENELLEPHVVTFTSDIYGVTFAGIQV
VSQTVSDYQDLQLEHRYALLQKHLLPAIAVPSVLAALDLSEERGARIAELAELPTGKD"
complement(3795. .4529)
/gene="orfX"
complement(3795. .4529)
/gene="orfX"
/codon_start=1
/transl_table=11
/product="3-ketoacyl-ACP/CoA reductase"
/protein_id="CAA07629.1"
/db_xref="GI:3425862"
/db_xref="GOA:O86853"
/db_xref="UniProt/TREMBL:O86853"
/db_xref="UniProt/TREMBL:O86853"
/translation="WTGSGRGIGRIALRLADGALVAVHYGSSFAAARETVETIRSS
GGQALAIRAEIVGVGDAALYANFADMGEGFVPEFDDILVNNAGVSGSGRITVETEE
```

```
VFRLVAVNVRAPLFLVHGLKRLDGGRIINISSAATRRAPFESIGYAMTKGAVDTL
TLARLQSGRGITVNAVAPGFVETDMNARRQTFPEAAALAAANSVFNIRGRPDDIAD
VWAFASDDSRWITGVVDATGTL"
complement(4746. .6446)
/gene="orfZ"
complement(4746. .6446)
/gene="orfZ"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAA07630.1"
/db_xref="GI:3451475"
/db_xref="InterPro:IPR007527"
/db_xref="UniProt/TREMBL:O86854"
/translation="MWRVRCGLVGRRRRRRRPKASRPRGAAPARWSSPAARQORPA
GGGTADARARAVDFAAGAPFLPPEADQAEPLPAGEQAGTSAGGATNAHAPAPS
PGTAHAEPASPAGDQTQSVASVRGRTVSSAPAKARAPRSMAPDGDURRTFPAL
PPREAAEGFAATWGNAMVTALEEGDAARERGRGYAHGHVDAITVTPGLVLAY
VRSRSRPPYQVRLRLTGLSDMDRFLDAAVERPGHIAALIDGELPHSLADLRGVP
LLPGPGLAPRCSQPSDSGHPCKHAAALCYOTARLLDADPFVLLLRGRGERALLDALS
RRNARARAAQDRGPGPLPGVRAGAALTFRALPLPALPAPHPQPPAPYAPAGG
PYPPFLDHLTADAAARAHALLTGRDPVGGITLMQDARVRLAARPGSLTAGTALYA
SLASAGRDYAEALARAAVAAWRQGLAGLDVLEEDPPAGFRDARPMMLAADLPAFR
PWRNRLTHPRGHVQLRLGRTGLWYAYESELGREDWWRPRTDPLDPVGLTSLGGPGDP
"
ORIGIN
Query Match 100.0%; Score 4346; DB 1; Length 6727;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCACGACGGCTCGGGTTCCGACGCGACGGGTACTCGTTCCGGCCACCGGSCACCG 60
DB 772 GTCACGACGGCTCGGGTTCCGACGCGACGGGTACTCGTTCCGGCCACCGGSCACCG 831
QY 61 GGTCTGGCTCGATGACCGACCGCATCGAGGACGTGCGCTCGCGGCTCTCTGATAGTGAGC 120
DB 832 GGTCTGGCTCGATGACCGACCGCATCGAGGACGTGCGCTCGCGGCTCTCTGATAGTGAGC 891
QY 121 GGGCCCGCGCGGACGACACATCGAGTTCATCTCCACATGCGGCCCGCCGGAAGTGAAGC 180
DB 892 GGGCCCGCGCGGACGACACATCGAGTTCATCTCCACATGCGGCCCGCCGGAAGTGAAGC 951
QY 181 ACCGACCGCGGACGTGACGCCATGAGGAGGGCCACGTCGCGGACGGATCACCCCTGGCT 240
DB 952 ACCGACCGCGGACGTGACGCCATGAGGAGGGCCACGTCGCGGACGGATCACCCCTGGCT 1011
QY 241 TCGGCCGAAGCTTTCGGCTGGTTCGCGCCCGACAGATGCGGAAACGCGCTTGGCGGGCGCGCC 300
DB 1012 TCGGCCGAAGCTTTCGGCTGGTTCGCGCCCGACAGATGCGGAAACGCGCTTGGCGGGCGCGCC 1071
QY 301 CGTCACTTCCGCGACGGTCGCGACGACCTGCGCCCTTGGCCCGCCCGCCGCTGCGCTCGGC 360
DB 1072 CGTCACTTCCGCGACGGTCGCGACGACCTGCGCCCTTGGCCCGCCCGCCGCTGCGCTCGGC 1131
QY 361 GCTCTCCAGGAACGGTTCGCGACGGGCTCGGATCTTCCGAGCATCTGCTGCGCGCGC 420
DB 1132 GCTCTCCAGGAACGGTTCGCGACGGGCTCGGATCTTCCGAGCATCTGCTGCGCGCGC 1191
QY 421 CGCTCTCAGGCCCGACGCTCTCTCGAAACGACGTGACCGCCCGCCAGCACCTCGAGAGCGCGCGC 480
DB 1192 CGCTCTCAGGCCCGACGCTCTCTCGAAACGACGTGACCGCCCGCCAGCACCTCGAGAGCGCGCGC 1251
QY 481 CGTCTGCTCTGCGCGGTGATGCTGCTGCGGCCCGGACAGCGGTACGCCGCTCCCTCGTG 540
DB 1252 CGTCTGCTCTGCGCGGTGATGCTGCTGCGGCCCGGACAGCGGTACGCCGCTCCCTCGTG 1311
QY 541 GCGGGGCCGGTCACTGCGCTGACCGCCACTTCCGCGATTCGCGCGGATCGACGACGCGC 600
DB 1312 GCGGGGCCGGTCACTGCGCTGACCGCCACTTCCGCGATTCGCGCGGATCGACGACGCGC 1371
QY 601 AACCGGGACGCTGCGGTACAGCGCGCGGACCAACGCGCTCGGACCGGATGGCGGGCGGCCA 660
```

Db 1372 AACGGGACGTCGCGTACAGCGCGCGGACACAGCGCGTCGGACCGGATGGCGGGCGCCCA 1431  
Qy 661 GGACACGCGTGTGGACATGAAAGTCTCTGGCCCGCAGGAAGTCCAGTCTAGCCCGGACATC 720  
Db 1432 GGAACGCGTGTGGACATGAAAGTCTCTGGCCCGCAGGAAGTCCAGTCTAGCCCGGACATC 1491  
Qy 721 GGGTACGGCCCGCTCGTCTCGGCTGCGCGCGCGTGAAGTCTGCGCGCGCGGTTTC 780  
Db 1492 GCGTACGGCCCGCTCGTCTCGGCTGCGCGCGCGTGAAGTCTGCGCGCGCGGTTTC 1551  
Qy 781 CCCACCGCGACATGGAACAGTCTCACACAGGTGCGCGACCGCGCTCGCGCGCGCGCGC 840  
Db 1552 CCCACCGCGACATGGAACAGTCTCACACAGGTGCGCGACCGCGCTCGCGCGCGCGCGC 1611  
Qy 841 CCGGAAACGCTCGTCTCGGCTCGGTCAGTCTGTCGACAGGAAACCGCGCGCGAC 900  
Db 1612 CCGGAAACGCTCGTCTCGGCTCGGTCAGTCTGTCGACAGGAAACCGCGCGCGAC 1671  
Qy 901 CCGTGTGAGGCGCGCTCCAGGCGCGGCGGTCGCGGTACTCGCGCGCGACACTCGAC 960  
Db 1672 CCGTGTGAGGCGCGCTCCAGGCGCGGCGGTCGCGGTACTCGCGCGCGACACTCGAC 1731  
Qy 961 GTTCCGGCGCGTACAGGTCACCGCTTCCGGCGCGCGCGGAGGACTCTGACGGGACCGGT 1020  
Db 1732 GTTCCGGCGCGTACAGGTCACCGCTTCCGGCGCGCGCGGAGGACTCTGACGGGACCGGT 1791  
Qy 1021 CCGGCGCAGCAGGTGGGCGACCTGACGCGCGCACACACCGGTACGCGCGGTCAAGAAT 1080  
Db 1792 CCGGCGCAGCAGGTGGGCGACCTGACGCGCGCACACACCGGTACACCGGTCAAGAAT 1851  
Qy 1081 CACTCGGGCTCTCTCGGGCAGCAGGAGGAGGGGCGCTCCGAAATATATAGAGGGA 1140  
Db 1852 CACTCGGGCTCTCTCGGGCAGCAGGAGGAGGGGCGCTCCGAAATATATAGAGGGA 1911  
Qy 1141 AGGGCAGGATCTGCCCCCGGGCGGAAACCGGCGATGTTCCGCGCCCCGGGCGGTTC 1200  
Db 1912 AGGGCAGGATCTGCCCCCGGGCGGAAACCGGCGATGTTCCGCGCCCCGGGCGGTTC 1971  
Qy 1201 AGCGGAGAAACCGGGCGCGCAGCGTGTGAGGAAGACGAGGTGCGCTCTGATGCC 1260  
Db 1972 AGCGGAGAAACCGGGCGCGCAGCGTGTGAGGAAGACGAGGTGCGCTCTGATGCC 2031  
Qy 1261 CCGTGAACCGCAGCGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320  
Db 2032 CCGTGAACCGCAGCGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2091  
Qy 1321 AGCAGCGGTGTGAACTCCGCGTACCGGACGAAACCGGCGTCCGCGATCGACGCGACGA 1380  
Db 2092 AGCAGCGGTGTGAACTCCGCGTACCGGACGAAACCGGCGTCCGCGATCGACGCGACGA 2151  
Qy 1381 AGGGCGCGCGGACCGGTACAGGCGACCGCGCTCGCGCTCGCGCTCGAGCAGCATGC 1440  
Db 2152 AGGGCGCGCGGACCGGTACAGGCGACCGCGCTCGCGCTCGCGCTCGAGCAGCATGC 2211  
Qy 1441 CCGGTAGTGTGTGTTGGGGCGGTGAAAGAGGGTGGGGTGTGATGTGTCTCACCCGCGATC 1500  
Db 2212 CCGGTAGTGTGTGTTGGGGCGGTGAAAGAGGGTGGGGTGTGATGTGTCTCACCCGCGATC 2271  
Qy 1501 GCGAGTGTCTGCTCGGAACTCGCGCAGGACGACGCTCTCTGTCGCGGGTGGCGCGG 1560  
Db 2272 GCGAGTGTCTGCTCGGAACTCGCGCAGGACGACGCTCTCTGTCGCGGGTGGCGCGG 2331  
Qy 1561 CCGCGCGCGCGGACCGGCGCGGTCCCGGGCACCGATCGGCTGGGAGTTCGCGAAAGTCGC 1620  
Db 2332 CCGCGCGCGCGGACCGGCGCGGTCCCGGGCACCGATCGGCTGGGAGTTCGCGAAAGTCGC 2391  
Qy 1621 CCGCATCCCGGCTAGACTTGAGGACTGTGAGCGCGTGTGAGCGGTGCGGAGTCCCGGTGGCAG 1680  
Db 2392 CCGCATCCCGGCTAGACTTGAGGACTGTGAGCGCGTGTGAGCGGTGCGGAGTCCCGGTGGCAG 2451  
Qy 1681 CAGCGCTCCGGCGCGGCGACCGCGCGAGTCACTGCTCTGTCAGGCGGTGCGCGCGCGC 1740  
Db 2452 CAGCGCTCCGGCGCGGCGACCGCGCGAGTCACTGCTCTGTCAGGCGGTGCGCGCGCGC 2511

Qy 1741 GGAATTCAGCTGGGAACAGGCCACTTCCACTCCAGCTCCGCGACCTCGCCCGACAGC 1800  
Db 2512 GGAATTCAGCTGGGAACAGGCCACTTCCACTCCAGCTCCGCGACCTCGCCCGACAGC 2571  
Qy 1801 CGAGTGTGTCCAGGTGGCAGGTGTAGTCCAGCGTGGCCATCAGGAAGTGTAGCCACCG 1860  
Db 2572 CGAGTGTGTCCAGGTGGCAGGTGTAGTCCAGCGTGGCCATCAGGAAGTGTAGCCACCG 2631  
Qy 1861 GCAGCCGTAGCCGCGTGGGAAGACAGGATCGCGCTCGACGCGAGGTCTCGGCGATCA 1920  
Db 2632 GCAGCCGTAGCCGCGTGGGAAGACAGGATCGCGCTCGACGCGAGGTCTCGGCGATCA 2691  
Qy 1921 GCAGCGGATCGTGTGCGTCCCGTGGACCGGTGCGAAGAACGGGTGCTGCGGGCAGGA 1980  
Db 2692 GCAGCGGATCGTGTGCGTCCCGTGGACCGGTGCGAAGAACGGGTGCTGCGGGCAGGA 2751  
Qy 1981 CCGCGGTGACCGAGAAACCGGTCCCGCTTCGCTATCCAGCTGACCGGGAACGGGTCT 2040  
Db 2752 CCGCGGTGACCGAGAAACCGGTCCCGCTTCGCTATCCAGCTGACCGGGAACGGGTCT 2811  
Qy 2041 GCACCTCGTCCGCTGGGCAAGCGCATCGGAACCGGCAATGCGGTGTTGTCGATCGAGT 2100  
Db 2812 GCACCTCGTCCGCTGGGCAAGCGCATCGGAACCGGCAATGCGGTGTTGTCGATCGAGT 2871  
Qy 2101 TGGCATCGGACGCGAATTTGATCAAACTACTGCTTCGGGCAATGGGTCCCCCAGGAAT 2160  
Db 2872 TGGCATCGGACGCGAATTTGATCAAACTACTGCTTCGGGCAATGGGTCCCCCAGGAAT 2931  
Qy 2161 CATGTGATCGCGAGCTGTTCTGTATGCGCGAACTTAAAGATACAGACTGAGCGGTTTTT 2220  
Db 2932 CATGTGATCGCGAGCTGTTCTGTATGCGCGAACTTAAAGATACAGACTGAGCGGTTTTT 2991  
Qy 2221 TTCTATCTTCCCGGGGAGACATGAACAAAGAGGACAGGCAATGGCAAGCAGGACCGGCG 2280  
Db 2992 TTCTATCTTCCCGGGGAGACATGAACAAAGAGGACAGGCAATGGCAAGCAGGACCGGCG 3051  
Qy 2281 GATCCGACGCGGACGATCTCTGGAACCGCGCGCGCGAGGTCTTCGAGAACAGCGGCTA 2340  
Db 3052 GATCCGACGCGGACGATCTCTGGAACCGCGCGCGCGAGGTCTTCGAGAACAGCGGCTA 3111  
Qy 2341 CCAGCTGCGACGATCAGGAGATCTCAAGGTGCGCGGGTGACCAAGGAGCGCTCTTA 2400  
Db 3112 CCAGCTGCGACGATCAGGAGATCTCAAGGTGCGCGGGTGACCAAGGAGCGCTCTTA 3171  
Qy 2401 TTTCACCTTCCAGTCCAAAGGAAGAACTGGCGTGGCGGTTCGACGCGCAGGAACCA 2460  
Db 3172 TTTCACCTTCCAGTCCAAAGGAAGAACTGGCGTGGCGGTTCGACGCGCAGGAACCA 3231  
Qy 2461 ACAGGCGGTTCCGAGCAACCGCTCCGCTGCAAGAACTCATCTGACATGGGCAATGTGTT 2520  
Db 3232 ACAGGCGGTTCCGAGCAACCGCTCCGCTGCAAGAACTCATCTGACATGGGCAATGTGTT 3291  
Qy 2521 CTGTCAACCGTTCGCGACGAACTGCGCGCGCGCGCGCGCGCTCTCCATGGACCA 2580  
Db 3292 CTGTCAACCGTTCGCGACGAACTGCGCGCGCGCGCGCGCGCTCTCCATGGACCA 3351  
Qy 2581 GCAGCGCACGCTCTCGATCGCGGAGAACCTTCCTCGTGGCGACGAGACACTCTCTGAA 2640  
Db 3352 GCAGCGCACGCTCTCGATCGCGGAGAACCTTCCTCGTGGCGACGAGACACTCTCTGAA 3411  
Qy 2641 GCTGTGAACCAAGGCGCAAGGAGAAAGGTGAGTGTGCTGCCCATGTGGTCAACCGGCTC 2700  
Db 3412 GCTGTGAACCAAGGCGCAAGGAGAAAGGTGAGTGTGCTGCCCATGTGGTCAACCGGCTC 3471  
Qy 2701 GCGCGATCTCTACGTGGGCGAGTTCGCGGGGATACAGGTCTGCTCCAGACGCTCAGCGA 2760  
Db 3472 GCGCGATCTCTACGTGGGCGAGTTCGCGGGGATACAGGTCTGCTCCAGACGCTCAGCGA 3531  
Qy 2761 CTACGAGGACCTCGAAACCGCTAGCGCTGCTGAGGAAGACATCTCTGCGCGGCTCAGC 2820  
Db 3532 CTACGAGGACCTCGAAACCGCTAGCGCTGCTGAGGAAGACATCTCTGCGCGGCTCAGC 3591



Qy	2821	GGTTCCCTCCGTGCTGGCCGCGCTCGATCTCTCCGAGAGCGCGGAGCAGCGCTCGGGC	2880
Db	3592	GGTTCCCTCCGTGCTGGCCGCGCTCGATCTCTCCGAGAGCGCGGAGCAGCGCTCGGGC	3651
Qy	2881	CGAATGSCACCGACCGGGGAAGGACTGACCCCGAAGCGCCCGCACCGGATACCGAACCG	2940
Db	3652	CGAATGSCACCGACCGGGGAAGGACTGACCCCGAAGCGCCCGCACCGGATACCGAACCG	3711
Qy	2941	CCGTGCCGAGCGCGACCGGGGCGCTACGAGGCGCGGGCGGGCGGGCGGTAGGTCTGC	3000
Db	3712	CCGTGCCGAGCGCGACCGGGGCGCTTACGAGGCGCGGGCGGGCGGGCGGTAGGTCTGC	3771
Qy	3001	CCTCGTACCGAAGCGTGGCGGCTCAGAGAACTGTTCCGCTGTGGCACTCAGCTACTGG	3060
Db	3772	CCTCGTACCGAAGCGTGGCGGCTCAGAGAACTGTTCCGCTGTGGCACTCAGCTACTGG	3831
Qy	3061	CCGCTGATCCACCGTGAAGTCCGTCCGAGGCGCAGAAAGGCCACCAAGTCCGCGATGCTGCG	3120
Db	3832	CCGCTGATCCACCGTGAAGTCCGTCCGAGGCGCAGAAAGGCCACCAAGTCCGCGATGCTGCG	3891
Qy	3121	GGTTCGCGATGCGGTTGAAACAGGAGTTGGGGCGAGTGGCGCGCGCTCGGGGGTC	3180
Db	3892	GGTTCGCGATGCGGTTGAAACAGGAGTTGGGGCGAGTGGCGCGCGCTCGGGGGTC	3951
Qy	3181	TGCGCGCGCGTGGTTCATGTCGCTCCACGAAACCCGGGCGCACCGCGTTGACCGTG	3240
Db	3952	TGCGCGCGCGTGGTTCATGTCGCTCCACGAAACCCGGGCGCACCGCGTTGACCGTG	4011
Qy	3241	ATCCCCCGTCCCCCAGTTGCTTGCCACGAGCGAGCGTGAGCGGTGTCACCGCACCTTG	3300
Db	4012	ATCCCCCGTCCCCCAGTTGCTTGCCACGAGCGAGCGTGAGCGGTGTCACCGCACCTTG	4071
Qy	3301	GTCATCGCGTATCCGATCGAATCGGGGAAACCGCGCGCGGGTCGCGGCGACGAGATGTTG	3360
Db	4072	GTCATCGCGTATCCGATCGAATCGGGGAAACCGCGCGCGGGTCGCGGCGACGAGATGTTG	4131
Qy	3361	ATGATCCGCGCGCGTCCGCGAGTGGTTTCAATCGCTCGTCCGACGAGAACGCGGTGCC	3420
Db	4132	ATGATCCGCGCGCGTCCGCGAGTGGTTTCAATCGCTCGTCCGACGAGAACGCGGTGCC	4191
Qy	3421	CGGAGTTGACGGCGACGAGTCCGTCGAGAACCTCTCGGTGACTTCCGTTGATCCGTCCC	3480
Db	4192	CGGAGTTGACGGCGACGAGTCCGTCGAGAACCTCTCGGTGACTTCCGTTGATCCGTCCC	4251
Qy	3481	GAGCGCTGACCGCGCGTGTGTTACAGAGATGTCGAACTCGGGCGGCACTCCGAACTCG	3540
Db	4252	GAGCGCTGACCGCGCGTGTGTTACAGAGATGTCGAACTCGGGCGGCACTCCGAACTCG	4311
Qy	3541	CCATCCCGGCGTGAACCGCGCGTAGAGCGGGCGCGGTACCCACGACCGGAGTTTCG	3600
Db	4312	CCATCCCGGCGTGAACCGCGCGTAGAGCGGGCGCGGTACCCACGACCGGAGTTTCG	4371
Qy	3601	GCCGGATGGCCCAACGCTGTCCGCGCTGTCCGGATGGTTCGACGGTCTCTCGCGCC	3660
Db	4372	GCCGGATGGCCCAACGCTGTCCGCGCTGTCCGGATGGTTCGACGGTCTCTCGCGCC	4431
Qy	3661	GCCGCTCGCTGCTCCGCTAGTGACTGCGACGAGCGCCCGGTCCGCGGCCAGCGCAGG	3720
Db	4432	GCCGCTCGCTGCTCCGCTAGTGACTGCGACGAGCGCCCGGTCCGCGGCCAGCGCAGG	4491
Qy	3721	CGGATACCGGTCGATGCGCGCTTCCCGGTCACGAGGCGGTCTTGCCCTCCAGC	3780
Db	4492	CGGATACCGGTCGATGCGCGCTTCCCGGTCACGAGGCGGTCTTGCCCTCCAGC	4551
Qy	3781	GGTCTTCCATACCTCGTCCCATGTCAGCATATCAGCCCGCGCTCGTGAGGAGCC	3840
Db	4552	GGTCTTCCATACCTCGTCCCATGTCAGCATATCAGCCCGCGCTCGTGAGGAGCC	4611
Qy	3841	ATGGCGCGCGTTCGGCGGTTTGAATCGAGGCTCAGAGCTTACCTGTGACCGCGTCAGACG	3900
Db	4612	ATGGCGCGCGTTCGGCGGTTTGAATCGAGGCTCAGAGCTTACCTGTGACCGCGTCAGACG	4671
Qy	3901	GGGCGCGAGTGCCCGGTTTGGACCGCTGGGGCCAGATCGGGGCGCGCACGGGGAAACCG	3960

Db	4672	GGGCGCGAGTGGCCCGGTTGGAACGGCTGGGGCCAGATCGGGCGCGCGCACGGGGAAACCG	4731
Qy	3961	CGCGCGGTCAAGGGTCAAGGGTCCCGGGACCGCCAGGCGGGTCAAGGGCACCGACCGGA	4020
Db	4732	CGCGCGGTCAAGGGTCAAGGGTCCCGGGACCGCCAGGCGGGTCAAGGGCACCGACCGGA	4791
Qy	4021	TCGAGTTCGCGGTCGCCACCGCGCCACGCTCTTCGCGGGCCAGCTCCGACTCCGACTCGTAGCGG	4080
Db	4792	TCGAGTTCGCGGTCGCCACCGCGCCACGCTCTTCGCGGGCCAGCTCCGACTCCGACTCGTAGCGG	4851
Qy	4081	TACAGAGCCCGGTTCGCGCCGAGTCTGAGCTTGAGCTGGGCGCGCGGGTGGGTGAGCGG	4140
Db	4852	TACAGAGCCCGGTTCGCGCCGAGTCTGAGCTTGAGCTGGGCGCGCGGGTGGGTGAGCGG	4911
Qy	4141	TTGCGCCAGGGGCGGAAAGGGGGAGGTTCGGCGCGGAGCATCATGGGGCGGGCGGGTTCG	4200
Db	4912	TTGCGCCAGGGGCGGAAAGGGGGAGGTTCGGCGCGGAGCATCATGGGGCGGGCGGGTTCG	4971
Qy	4201	AAACGGCGCGCGGGTCCAGGGCTCTTCAGAGACGCTTAGACCCGCCAACCCGCC	4260
Db	4972	AAACGGCGCGCGGGTCCAGGGCTCTTCAGAGACGCTTAGACCCGCCAACCCGCC	5031
Qy	4261	TGCCCGCAGGCGGCGAACCGCGCCGCGCAGCTCCGCGCTGTGCTCCGGCGCGGCGAGGCG	4320
Db	5032	TGCCCGCAGGCGGCGAACCGCGCCGCGCAGCTCCGCGCTGTGCTCCGGCGCGGCGAGGCG	5091
Qy	4321	AGCAGCGGTAGAGCGCGCGGTAC	4346
Db	5092	AGCAGCGGTAGAGCGCGCGGTAC	5117

RESULT 2  
SCO939127/c

LOCUS 290850 bp DNA linear BCT 16-APR-2005  
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 24/29.  
ACCESSION AL039127 AL023496 AL023861 AL031182 AL031232 AL035161  
AL035205 AL035206 AL132824 AL512902 AL590982 AL591083 AL591084  
AL645882  
AL939127.1 GI:24429552

## KEYWORDS

Streptomyces coelicolor A3(2)

## SOURCE

Streptomyces coelicolor A3(2)

## ORGANISM

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

## REFERENCE

1

## AUTHORS

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornaby, T., Howarth, S., Huang, C.H., Kiebert, T., Larke, D., Murphy, L., Oliver, K., O'Neill, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorek, A., Woodward, J., Barrell, B.G., Parkhill, J. and Hopwood, D.A.

Complete genome sequence of the model actinomycete Streptomyces

coelicolor A3(2)

Nature 417 (6885), 141-147 (2002)

12000953

2 (bases 1 to 290850)

Bentley, S.D.

Direct Submission

Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk

On or before Oct 30, 2002 this sequence version replaced

gi:20520661, gi:20520749, gi:20520754, gi:20520755, gi:20520688, gi:20520689, gi:20520815, gi:20520683, gi:20520917, gi:20520686, gi:20520864.

Location/Qualifiers

1. 290850

/organism="Streptomyces coelicolor A3(2)"

/mol\_type="genomic DNA"

```

/strain="A3(2)"
/db_xref="taxon:100226"
97. .1101
/gene="SCO6257"
/notes="synonym: SCAH10.22"
97. .1101
/gene="SCO6258"
/notes="SCAH10.22, probable ABC transport system sugar binding lipoprotein, len: 334 aa; similar to TR:CA841563 (EMBL:AL049727) Streptomyces coelicolor putative secreted solute binding protein, 337 aa; fasta scores: opt: 1003 z-score: 1130.9 E(): 0; 49.8% identity in 325 aa overlap, to SW:R8SB_BACSU (EMBL:Z92953) Bacillus subtilis D-ribose-binding protein precursor Rb8B, 305 aa; fasta scores: opt: 206 z-score: 238.7 E(): 6.3e-06; 23.0% identity in 235 aa overlap and to TR:CA841563 (EMBL:SC9B1) Streptomyces coelicolor SC9B1.17c, 337 aa; fasta scores: opt: 1003 z-score: 1045.2 E(): 0; 49.8% identity in 325 aa overlap. Contains match to Pfam entry PF00532 Peripla_BP_like, Periplasmic binding proteins and LacI family and a match situated in the correct position to Prosite entry PS00013 Prokaryotic membrane lipoprotein lipid attachment site. Contains also possible N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative ABC transport system sugar binding lipoprotein"
/protein_id="CAB60176.1"
/db_xref="GI:6273662"
/db_xref="InterPro:IPR001761"
/db_xref="UniProt/TREMBL:Q9RKT5"
/translation="MARFTWGIAGALGSVSLAGSCSTGGKRAEDARKAAABEGR AADTPRTFAMITHSQDGTWDIVQSGAEQAAVDNINFLYSHDAEQQAQLVNA AIDKVDGIIVTLAKDPKMSALAKAHKIGIPVITVNSGSESKERFALTHVGDRTIA GRAVEHLENERQAVQVILHEQGNVGHQSCDGVKTFDQKQVLYNGTSMWPVQS AIEAKLQTDKSDVAVTLCVPAHYADTAVKAKQAGSKAEIDTFDLNAKVAAGLADGTLG FAVDQOPLYQGEAVDILLKYKNADVLGGSPVLTPGQIITKDAAALADYTKRGT R"
133. .165
/misc_feature
/gene="SCO6257"
/notes="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
241. .966
/misc_feature
/gene="SCO6257"
/notes="Pfam match to entry PF00532 Peripla_BP_like, periplasmic binding proteins and LacI family., score 13.80, E-value 2.1e-06"
1098. .2138
/gene="SCO6258"
/notes="synonym: SCAH10.23"
1098. .2138
/gene="SCO6258"
/notes="SCAH10.23, probable ABC transport system sugar permease, len: 346 aa; similar to TR:O68120 (EMBL:AF010496) Rhodospirillum rubrum capsule protein RbSC, 345 aa; fasta scores: opt: 445 z-score: 508.0 E(): 6.4e-21; 30.5% identity in 351 aa overlap and to SW:R8SC_ECOLI (EMBL:L10328) Escherichia coli ribose transport system permease protein RbSC, len: 321 aa; fasta scores: opt: 286 z-score: 329.9 E(): 5.3e-11; 30.1% identity in 329 aa overlap. Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative ABC transport system sugar permease"
/protein_id="CAB60177.1"
/db_xref="GI:6273663"
/db_xref="GOA:Q9RKT4"
/db_xref="InterPro:IPR001851"
/db_xref="UniProt/TREMBL:Q9RKT4"
/translation="MNATQVKAADRIQLTQRKLLGRPELGSVVGATVAVFFFAFF ADSFLHAASLSTLVYAASTIGINAVPVALMIGEDFLSAGVNVTSALVSMFSYQM
TANVGVVSVLLVTLTAIGAFNGFMLTRTKLPSPFITLTGTFMLTGLNLGPTKLVVDGT
VSTKSIADMEGFPSPADQVFPASTITIGGVGFKVTLMLWALVAVASWILLRTRAGNWIF
AVGNGKDAARAVGVPPVAKTKIGLTMGVGFGAWISQHLHLSFYDVSQSEGSGVNEIYI
IAVIVGGLITGGTGGSAVGAFAFGMTSGKIVFABWNPDPWFKFLGAMLLATLL
NAWVRKRAEATK"
2124. .2127
/gene="SCO6258"
2135. .2926
/gene="SCO6259"
/notes="synonym: SCAH10.24"
2135. .2926
/gene="SCO6259"
/notes="SCAH10.24, probable ABC sugar transport ATP binding protein, len: 263 aa; similar to TR:O68121 (EMBL:AF010496) Rhodospirillum rubrum capsule protein RbSC, 305 aa; fasta scores: opt: 612 z-score: 706.1 E(): 5.9e-32; 45.6% identity in 217 aa overlap, to SW:RBSA_ECOLI (EMBL:M13169) Escherichia coli, ribose transport ATP-binding protein RbsA, 501 aa; fasta scores: opt: 509 z-score: 585.6 E(): 3e-25; 31.5% identity in 254 aa overlap and to TR:CAAL5787 (EMBL:SC7B7) Streptomyces coelicolor St7B7.07, 260 aa; fasta scores: opt: 592 z-score: 521.2 E(): 1e-23; 43.1% identity in 246 aa overlap. Contains Pfam match to entry PF00005 ABC tran, ABC transporter and Prosite PS00017 ATP/GTP-binding site motif A (P-loop) and PS00211 ABC transporters family signature"
/codon_start=1
/transl_table=11
/product="probable ABC sugar transport ATP binding protein"
/protein_id="CAB60178.1"
/db_xref="GI:6273664"
/db_xref="GOA:Q9RKT3"
/db_xref="InterPro:IPR003439"
/db_xref="InterPro:IPR003593"
/db_xref="UniProt/TREMBL:Q9RKT3"
/translation="MTRNEDRTALVELSGVKNVGRALGVSLVHAGELTCVLGD NGAKSTLIKIAGLHQHGGTSLDGETRUSPREALDRGIATVYQLAVVPLMPV WRNFFLSEPRKGVAPFKRMDVDMRRTRTHAELRMGIDLRDVPDQPIGLSGGRCV IARAVYFGAKVLVDLEPTALGVQSGVLYKVAARDQGLGVVLIITHNPHAYLVG DRFVLKRGTMVANQTRDBVTLDLTQMGAGTDLDLRHELRG"
2240. .2815
/gene="SCO6259"
/notes="Pfam match to entry PF00005 ABC tran, ABC transporter, score 168.70, E-value 9.6e-47"
2261. .2284
/gene="SCO6259"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
2588. .2632
/gene="SCO6259"
/notes="PS00211 ABC transporters family signature"
3042. .4190
/gene="SCO6260"
/notes="synonym: SCAH10.25"
3042. .4190
/gene="SCO6260"
/notes="SCAH10.25, possible sugar kinase, len: 382 aa; similar to SW:GLK_STRCO (EMBL:X65932) Streptomyces coelicolor glucokinase (EC 2.7.1.2) (Glucose kinase) Glk, 317 aa; fasta scores: opt: 1022 z-score: 1142.9 E(): 0; 46.6% identity in 311 aa overlap and to TR:CA851974 (EMBL:SC6E10) Streptomyces coelicolor SC6E10.20c, 317 aa; fasta scores: opt: 1022 z-score: 1035.4 E(): 0; 46.6% identity in 311 aa overlap. Contains a match to Pfam entry PF00480 ROK_ROK family and Prosite entry PS01125 ROK family signature"
/codon_start=1
/transl_table=11
/product="putative sugar kinase"
/protein_id="CAB60179.1"
/db_xref="GI:6273665"
/db_xref="GOA:Q9RKT2"

```

		/db_xref="InterPro:IPR000600" /db_xref="InterPro:IPR004654" /db_xref="UniProt/TREMBL:Q9RKT2" /translation="MSTVRDFTAPIGRRAPELVLTGTRRRSHLTAPRVPTVGIDIG GTKWAGVVDADGNLILSKRLTETPKSKPRKVEDTIVELVDLSRDHVDHVGIGAA GWVDNRNVLFAPHLSWRNPLRIAGRLAVPLVDNDANTAAWAEWFRGAGRGED HLVMTLTGTGGAILLEDGOVKRGKYGVAGFEHMQVVPBGHRCPCPNRGCEWQYSSG NALVREARELAASPSVAYGIIIEHVKSIGDITGPMITELAREGDAMCVPELLQIQWL LGVGIANLAALDPSFCVIGGVSAADDLLIGPARDAPKROLTGRTGRPEARIVRAQL GPEAGWGAADLSLVARRFRRAKRRVRYERYVAERRESRESL" 3159..3704 /gene="SC06260" /note="Pfam match to entry PF00480 ROK, ROK family, score 215.50, E-value 5.4e-63" 3540..3623		misc_feature		misc_feature	
Query Match		98.9%; Score 4299.8; DB 1; Length 290850;					
Best Local Similarity		99.8%; Pred. No. 0;					
Matches 4337; Conservative		0; Mismatches 7; Indels 3; Gaps 3;					
Qy	1	GTGACGACGGCGTGGGTTTCGACCGCGACGCGGTACTCGTTCCGGGCCACCGGCACCG	60	Db	14234	GTGACGACGGCGTGGGTTTCGACCGCGGTACTCGTTCCGGGCCACCGGCACCG	14175
Qy	61	GGTCTGGCTCGATGACCGACCGCATCGACGAGCGTTCGGCTGGCGCTCTCTGATGTGAGC	120	Db	14174	GGTCTGGCTCGATGACCGACCGCATCGACGAGCGTTCGGCTGGCGCTCTCTGATGTGAGC	14115
Qy	121	GGCCCCGGCGGCGACGACATCGACGTTCATCTCCACTGCGCCCCCGGAAAGTGTAGC	180	Db	14114	GGCCCCGGCGGCGACGACATCGACGTTCATCTCCACTGCGCCCCCGGAAAGTGTAGC	14056
Qy	181	ACCGCACCGCGACGTGACGCCATGGAGGGCCACGTCCGCGGACGGATCAACCTGGCT	240	Db	14055	ACCGCACCGCGACGTGACGCCATGGAGGGCCACGTCCGCGGACGGATCAACCTGGCT	13996
Qy	241	TGGCGCGAAGGTTCCGGTGGTTCGGCGCCCGAGATGCGGAACGGCTTGGCGGCGCGCC	300	Db	13995	TGGCGCGAAGGTTCCGGTGGTTCGGCGCCCGAGATGCGGAACGGCTTGGCGGCGCGCC	13936
Qy	301	CGTCACTTCCGACCGGTCCGACGACCTTGCGCTTGCGCCCGCGCTGCGCTCGGC	360	Db	13935	CGTCACTTCCGACCGGTCCGACGACCTTGCGCTTGCGCCCGCGCTGCGCTCGGC	13876
Qy	361	GCTCTCAGGAACCGGTCCGGACGGGCTCGGATCTTCGAGACATCTGTCGCGCGC	420	Db	13875	GCTCTCAGGAACCGGTCCGGACGGGCTCGGATCTTCGAGACATCTGTCGCGCGC	13816
Qy	421	CGCCTCCAGCCCCAGCTCTCGAAGCGAGTGACCGCCCCAGCACTCGGAGCGCGCGC	480	Db	13815	CGCCTCCAGCCCCAGCTCTCGAAGCGAGTGACCGCCCCAGCACTCGGAGCGCGCGC	13756
Qy	481	CGTCTGCTGCTGGCGGTGATCGCTCGGGCCCGACAGCGGTACGCCCGCTCCCTCGTG	540	Db	13755	CGTCTGCTGCTGGCGGTGATCGCTCGGGCCCGACAGCGGTACGCCCGCTCCCTCGTG	13696
Qy	541	GCCGGCGCGGTGAGTGGCCCTGACCGCCATTTCGCGATGTGCGCGGATCGACAGGC	600	Db	13695	GCCGGCGCGGTGAGTGGCCCTGACCGCCATTTCGCGATGTGCGCGGATCGACAGGC	13636
Qy	601	AACCGGGACGTGCGGTACAGCGCGGGACACGCGCTCGGACCGGATGGCGGGGCCCA	660	Db	13635	AACCGGGACGTGCGGTACAGCGCGGGACACGCGCTCGGACCGGATGGCGGGGCCCA	13576
Qy	661	GGACACGTGTTGGACATGAAGGTCTGGCCCCGACAGGAGGTCCAGTCTAGCCCGGACTC	720	Db	13575	GGACACGTGTTGGACATGAAGGTCTGGCCCCGACAGGAGGTCCAGTCTAGCCCGGACTC	13516
Qy	721	GCGTACGGCCCCCTCGTTCTCGCGCTGCGCGCGCTGATGAAGTCTGCGCGCCCGGTTTC	780	Db	13515	GCGTACGGCCCCCTCGTTCTCGCGCTGCGCGCGCTGATGAAGTCTGCGCGCCCGGTTTC	13456
Qy	781	CCCCACCGGACATGGAACGTTTACACAGGTTCGGGACCGCGGCTCGCGCGCGCGCGC	840	Db	13456	CCCCACCGGACATGGAACGTTTACACAGGTTCGGGACCGCGGCTCGCGCGCGCGCGC	13396
Qy	841	CGCGAAACCGTCTGTCGTCGGCTCGGTGGCACTGTTCTGTGACGAGGAACCGCGCGCGCAC	900	Db	13395	CGCGAAACCGTCTGTCGTCGGCTCGGTGGCACTGTTCTGTGACGAGGAACCGCGCGCGCAC	13336
Qy	901	CCGTTTCAGGGCCCGGTCCAGGCCCGGGCGGTGCGGTACTTCGCCCGCGACAGACTCGAC	960	Db	13335	CCGTTTCAGGGCCCGGTCCAGGCCCGGGCGGTGCGGTACTTCGCCCGCGACAGACTCGAC	13276
Qy	961	GTTCGGGCGCGGTGACCGGTCAACCGTTTCGGGCGCGCGGCGGAGGACTCTGACCGGACCGGT	1020	Db	13275	GTTCGGGCGCGGTGACCGGTCAACCGTTTCGGGCGCGCGGCGGAGGACTCTGACCGGACCGGT	13216
Qy	1021	CCGGGCGACGAGTGGGGGACCTGACCGGCGCACACACCGGTGACGCGGCTCAAGAAAT	1080	Db	13215	CCGGGCGACGAGTGGGGGACCTGACCGGCGCACACACCGGTGACGCGGCTCAAGAAAT	13156
Qy	1081	CACTCGGGGCTCTCTCGGGCAGCGAGCGGGCGCTCCGAAATACATATGAGGGGA	1140	Db	13155	CACTCGGGGCTCTCTCGGGCAGCGAGCGGGCGCTCCGAAATACATATGAGGGGA	13096
Qy	1141	AGGCGACGATCT-GCCGCCCGGGCGCGAACCGCGGATGTTTCGCCCGCGGGCCCGGTGCTT	1199	Db	13095	AGGCGACGATCT-GCCGCCCGGGCGCGAACCGCGGATGTTTCGCCCGCGGGCCCGGTGCTT	13037
Qy	1200	CAGCCGAGAAACCGCGGCGCGGACAGCGTGTGTAGGAAGACGAGGCTGCGTCTCTGATGC	1259	Db	13036	CAGCCGAGAAACCGCGGCGCGGACAGCGTGTGTAGGAAGACGAGGCTGCGTCTCTGATGC	12977
Qy	1259	CAGCCGAGAAACCGCGGCGCGGACAGCGTGTGTAGGAAGACGAGGCTGCGTCTCTGATGC	1319	Db	12976	CAGCCGAGAAACCGCGGCGCGGACAGCGTGTGTAGGAAGACGAGGCTGCGTCTCTGATGC	12917
Qy	1319	CCGTTGACCGGCAAGGTGTCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1379	Db	12916	CCGTTGACCGGCAAGGTGTCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	12857
Qy	1379	CAGCAACCGGCTGTGCAACTCTCGCTACCGAGCAACCGGGTTCGCGCGGATCGAGCGCACG	1439	Db	12856	CAGCAACCGGCTGTGCAACTCTCGCTACCGAGCAACCGGGTTCGCGCGGATCGAGCGCACG	12797
Qy	1439	CCCGGTACGTGTTGGGGCGCTGGAAGAGGGTTCGGGTGACTGTTCTCAACCGCGCAGT	1499	Db	1440	CCCGGTACGTGTTGGGGCGCTGGAAGAGGGTTCGGGTGACTGTTCTCAACCGCGCAGT	1499
Qy	1499	CCCGGTACGTGTTGGGGCGCTGGAAGAGGGTTCGGGTGACTGTTCTCAACCGCGCAGT	1559	Db	12796	CCCGGTACGTGTTGGGGCGCTGGAAGAGGGTTCGGGTGACTGTTCTCAACCGCGCAGT	12737
Qy	1559	CGCCACGTGTCCTGTGTCGAACTCGCGCACAGGACCACTGTCCTCGTCGCGGTGCGACCG	1619	Db	1500	CGCCACGTGTCCTGTGTCGAACTCGCGCACAGGACCACTGTCCTCGTCGCGGTGCGACCG	1559
Qy	1619	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	1679	Db	12736	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	12677
Qy	1679	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	1739	Db	1560	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	1619
Qy	1739	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	1799	Db	12676	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	12617
Qy	1799	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	1859	Db	1620	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	1679
Qy	1859	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	1919	Db	12616	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	12557
Qy	1919	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	1979	Db	1680	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	1739
Qy	1979	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	2039	Db	12555	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	12497
Qy	2039	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	2099	Db	1740	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	1799
Qy	2099	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	2159	Db	12496	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	12437
Qy	2159	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	2219	Db	1800	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	1859
Qy	2219	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	2279	Db	12436	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	12377
Qy	2279	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	2339	Db	1860	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	1919
Qy	2339	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	2399	Db	1919	CGCGCGCGCGCGGACCGGGCGCGGTTCGCGGCAACCGGATCGGTGGGAGTTCGCGAGTTCG	1919

12376	GGCACGCCCTAGACCGCGGTGGGAAGACAGAGCATCGCCGGCTCTGACGACGAGGGTCTCGCGCATC	12311
1920	AGCACGCGATCGTGTCCGTCCCGCTGGACACGGTGCAGAAACCGGTGTGTGTGGGCGAG	1979
12316	AGCAGCGGATCGTGTCCGTCCCGCTGGACACGGTGCAGAAACCGGTGTGTGGGCGAG	12257
1980	ACGGCGGTGACCGAGAACCGGTCAACGGCCCTTCGGTATTCAGCTGACCGGGAACGGCTCC	2039
12256	ACGGCGGTGACCGAGAACCGGTCAACGGCCCTTCGGTATTCAGCTGACCGGGAACGGCTCC	12197
2040	TGCACCCCTCGTCCGGTGGACAAGCGCCATCGGAACCGGCAATGCGGTTGTTCATCGAG	2099
12196	TGCACCCCTCGTCCGGTGGACAAGCGCCATCGGAACCGGCAATGCGGTTGTTCATCGAG	12137
2100	TTGGCATCGGACGCAAGATTGATCAAAACTACTGCTTCGGGCATGGGTCCCCCCCAGGAA	2159
12136	TTGGCATCGGACGCAAGATTGATCAAAACTACTGCTTCGGGCATGGGTCCCCCCCAGGAA	12077
2160	TCAATGATGCCGAGCTGTTCGTATGCGCGAAAGCTTAAAGATACAGACTGACGGTTTTT	2219
12076	TCAATGATGCCGAGCTGTTCGTATGCGCGAAAGCTTAAAGATACAGACTGACGGTTTTT	12017
2220	TTTTCTATCCTTCCCGGGGAGACATGAAACAAGAGCGAGGCATGSCCAAGCAGACCCGG	2279
12016	TTTTCTATCCTTCCCGGGGAGACATGAAACAAGAGCGAGGCATGSCCAAGCAGACCCGG	11957
2280	CGATCCGACGCGGACAGACGATCCTTGGACCGCCGGCGCAGGTCTTCAGAAAGCAGGGCT	2339
11956	CGATCCGACGCGGACAGACGATCCTTGGACCGCCGGCGCAGGTCTTCAGAAAGCAGGGCT	11897
2340	ACCAAGCTGCCAGATCAACGAGATCCTCAAGGTGGCCGGGTGACAAAGGGAGCCCTCT	2399
11896	ACCAAGCTGCCAGATCAACGAGATCCTCAAGGTGGCCGGGTGACAAAGGGAGCCCTCT	11837
2400	ACTTCACTTCCAGTCCAGGAGAACTGGCCGTGGGGGTCTTGACGCCCCAGAGAACAC	2459
11836	ACTTCACTTCCAGTCCAGGAGAACTGGCCGTGGGGGTCTTGACGCCCCAGAGAACAC	11777
2460	CACAGCCGTTCCGGAGCAACCCCTCCGGCTGCAGAAACTCATGACATGGGCATGTGT	2519
11776	CACAGCCGTTCCGGAGCAACCCCTCCGGCTGCAGAAACTCATGACATGGGCATGTGT	11717
2520	TCTGTCAACCGTTTGCAGCAAGACGTGTGGCCGGGCCGGCGTGGCTCTCCATGGACC	2579
11716	TCTGTCAACCGTTTGCAGCAAGACGTGTGGCCGGGCCGGCGTGGCTCTCCATGGACC	11657
2580	AGCAGGGCAGCGTCTCGATCGCCGAGGACCTTCGGTCGTGGCACGAGACATCTCTGA	2639
11656	AGCAGGGCAGCGTCTCGATCGCCGAGGACCTTCGGTCGTGGCACGAGACATCTCTGA	11597
2640	AGCTGTGAACCAAGGCCAAGGAGAACGGTGTGCTGCCCATGTGTGTACCAACCGACT	2699
11596	AGCTGTGAACCAAGGCCAAGGAGAACGGTGTGCTGCCCATGTGTGTACCAACCGACT	11537
2700	CGGCGCATCTTACGTGGGACGTTCCGCCGGGATACAGGTGTGTTCAGACGCTCAGCG	2759
11536	CGGCGCATCTTACGTGGGACGTTCCGCCGGGATACAGGTGTGTTCAGACGCTCAGCG	11477
2760	ACTTACGAGACTCGAACACCGCTACCGCTCTGCAGAGACACATCTCTGCCGCCATCG	2819
11476	ACTTACGAGACTCGAACACCGCTACCGCTCTGCAGAGACACATCTCTGCCGCCATCG	11417
2820	CGGTTCCCTCCGTGCTGGCCGCTCATGCTCTCCGAGGAGCGGAGACGCTCTCGGG	2879
11416	CGGTTCCCTCCGTGCTGGCCGCTCATGCTCTCCGAGGAGCGGAGACGCTCTCGGG	11357
2880	CCGAACTGGCACCGAACCGGGAAGGACTGACCGCCGAAGCCCGCACCGGATACCGACC	2939
11356	CCGAACTGGCACCGAACCGGGAAGGACTGACCGCCGAAGCCCGCACCGGATACCGACC	11297
2940	GCGTGTCCGAGCGGCGGACCGGGGCGGCTTACGGGCCCGGGCGGGCGGTAGGCTGTG	2999
11296	GCGTGTCCGAGCGGCGGACCGGGGCGGCTTACGGGCCCGGGCGGGCGGTAGGCTGTG	11237

QY 4080 GTACCAGACCGCGTCCGCGCCGAGTCTGAGCTGACGTGGCGCGCGCGGTGGGTGAGCG 4139  
Db |||||  
QY 10156 GTACCAGACCGCGTCCGCGCCGAGTCTGAGCTGACGTGGCGCGCGCGGTGGGTGAGCG 10097  
Db |||||  
QY 4140 GTTGGCGCAGCGCGGAGGAGCGGAGGTGCGCGCGGAGCAATCATGGCGCGCGCGGTCC 4199  
Db |||||  
QY 10096 GTTGGCGCAGCGCGGAGGAGCGGAGGTGCGCGCGGAGCAGCAGGCGCGCGCGGTCC 10037  
QY 4200 GAAACGGCGCGCGCGGTCCAGGGCTCTCCAGGAGCTCTAGACCGCGCAACCGGCC 4259  
Db |||||  
QY 10036 GAAACGGCGCGCGCGGTCCAGGGCTCTCCAGGAGCTCTAGACCGCGCAACCGGCC 9977  
QY 4260 CTGCGCGCAGCGCGGAGCGCGCGCGCGAGCTCCGCGGTGTCGGTCCGCGCGCGAGGC 4319  
Db |||||  
QY 9976 CTGCGCGCAGCGCGGAGCGCGCGCGCGAGCTCCGCGGTGTCGGTCCGCGCGCGAGGC 9917  
QY 4320 GAGCGACGCTAGAGCGCGCGGTACC 4346  
Db |||||  
QY 9916 GAGCGACGCTAGAGCGCGCGGTACC 9890  
Db |||||  
RESULT 3  
AB011413/c  
LOCUS  
DEFINITION Streptomyces griseus genes for Orf2, Orf3, Orf4, Orf5, AfsA, Orf8,  
partial and complete cds.  
ACCESSION AB011413  
VERSION AB011413.1 GI:3401946  
KEYWORDS Orf8; AfsA; Orf5; Orf4; Orf3; Orf2.  
SOURCE Streptomyces griseus  
ORGANISM Streptomyces griseus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomycetes.  
REFERENCE 1 (sites)  
AUTHORS Umeyama,T.  
TITLE Open reading frame encoded around afsA gene  
JOURNAL Unpublished  
2 (bases 1 to 12070)  
AUTHORS Umeyama,T.  
DIRECT SUBMISSION  
TITLE Submitted (23-FEB-1998) Takashi Umeyama, University of Tokyo,  
Department of Agriculture and Life Sciences, Yayoi 1-1-1,  
Bunkyo-ku, Tokyo 113, Japan  
JOURNAL (E-mail:aa67103@nongo.ecc.u-tokyo.ac.jp, Tel:++81-3-3812-2111)  
FEATURES  
source  
1..12070  
/organism="Streptomyces griseus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1911"  
complement(2660..3511)  
/gene="orf2"  
complement(2660..3511)  
/gene="orf2"  
/function="sensor histidine kinase"  
/codon\_start=1  
/transl\_table=11  
/product="Orf2"  
/protein\_id="BAA32130.1"  
/db\_xref="GI:3401947"  
/translation="VAAVPRAHPGPEAAAPARALTAAVTLITVUGLACLSMLVIR  
TDDRMDAYDEMRRAVSANSLVYEDDIRSAGPAHRSGLVSLVIGAGALTALA  
ALTGVLGRSGCAWEALQEORULADAEHLRTPVAVMRGSGVEAAGAGGLGQL  
PRIIRAADMDADVNLTRGLEAGTSGYRSLRLDOLVEEVCAELPEGGPGLES  
LEESVEADAALVRVAENLLDNAVRHGPSRGAAGRTAGDRVRTGVRVADRGPV  
APARLPS"  
complement(3523..4197)  
/function="two-component regulator protein"  
/codon\_start=1  
/transl\_table=11  
/product="Orf3"  
/protein\_id="BAA32131.1"  
/db\_xref="GI:3401948"  
/translation="NRVLLVEDDDLRDVIGNGLRDGGFAWDCASDWPEADVILLHLSA

YDCVLDRMVPSGDTLAPLEGRRRAGNSVPVLCITLALDSLRLRGLSGADDYLAKP  
LSVRLRVRLRGLSRASARLPFLGCADVMDVARHEVRGGVLLSLSPKEAVLQQ  
FVLRDVTVTRTGLLEHCWDEMADPVSNVVDVAVAGLRRLKLGSPGLVHTVRGQGFLLS  
ADPGPS"  
complement(4333..5250)  
/function="methyltransferase"  
/codon\_start=1  
/transl\_table=11  
/product="Orf4"  
/protein\_id="BAA32132.1"  
/db\_xref="GI:3401949"  
/translation="MATQALGTALRLGVFDRIGTGELRAEALAGSLGTHPRATLRLR  
ALAGLQLLSPPECAFRTAAGNVLRADAGTVMVARMFTDPVMLRGWLLDSVRT  
GETTFDVFQDFGLRHEPFLSAANEAMSGQRTLTAEVPHHYFGFRQRLVDIG  
GGTFTVLAJLRAHQEPGLVFDTAEGLAQAPRELAREGLDGRVTLTGDFFAAPAG  
GDLYLNSIIHDWDDAVRGHRLHRIDVIPHGSLLIIVEPVLPAVFPADRPDNLVSDL  
NMLVNVGGRERTADDFAAALCTAGGFACGA"  
complement(5450..6361)  
/function="regulator protein"  
/codon\_start=1  
/transl\_table=11  
/product="Orf5"  
/protein\_id="BAA32133.1"  
/db\_xref="GI:3401950"  
/translation="MSBLPLDQVRLTLLAVVDEGTFDSAAATLRLTPSAVSQRVKALEQ  
RTGRVLLMRTPKVRPTESGEVVALARLARLEHQAALGMSGPGEPTLLPAINVSDS  
LATSWQPCGACRSWDSATTPGCGPYGRPAAGAGDGRDLVAGGDLGFGPAAG  
RMRYLPVSPGFADRWLRDGLDRLRELIAGEAPVPCFDRDDLQDAFVRLGPGARPS  
ARRHLVPTSEGFANAVAGMGWGPVEQAEPLLSDGLRVLKLAPEPTVDVPLYWQWK  
LESPALAAVAEAAVAAAEALDEAPPN"  
7486..8391  
/codon\_start=1  
/transl\_table=11  
/product="AfsA"  
/protein\_id="BAA32134.1"  
/db\_xref="GI:3401951"  
/translation="MDAEVVPVGIEMVHRTPEDAFPENWVLRDRFAVBAVL  
HDHPPFAPVGDLDHPLVAAWMAQAALAFHAGYGLPVHFLLTLDYVCHPEHLG  
VGSEPTGILEVCSDLKWRAGLPAQGRVGMVHGRDLAATGAATVAATFSPKAVRRM  
RGDVFGEISLPTETAPVSPAGRVEDVVLSTGREGVWELVDRTHPTLFORPND  
HVPQMLLEAARQAACLVAGPAGIVPVEARTFRHRYSEFGSPCMVAGVAVQGADEDTV  
TVPTGHQDGETVFSTVLSGPRAGH"  
10558..11598  
/function="alcohol dehydrogenase"  
/codon\_start=1  
/transl\_table=11  
/product="Orf8"  
/protein\_id="BAA32135.1"  
/db\_xref="GI:3401952"  
/translation="WTTVAAYAPRAKAPLERTTIERRPVGEFDILIDIKFAGICHSD  
IHQARDWGEGIFPMVPGHEIAGIVAGTGPVTRFKVGDVGVGCMVDSCGTCDACLM  
GREQHCAGNTQYNALDRNGEPTVGGYSTHLVTEKPAFALSIPIGIALDEAAPLLCAG  
ITTSPLRRMGAGPKKVAVVGLGGLGHMAVIAHAGAEVTVLSQSRKKKDDGLKLG  
ADHYATGDPRTPELAGTAVILSTVAPLDGFAYLALLTEGLTVNVGAPPEPVS  
NLFSLLGNRSIAGSAGIIBETQMLDFCAVHGLGAEIEVIGAQVNEAVERVLASD  
VRYRPFVIDTATI"  
Query Match 20.7%; Score 901.2; DB 1; Length 12070;  
Best Local Similarity 68.6%; Pred. No. 6e-69;  
Matches 1361; Conservative 0; Mismatches 508; Indels 115; Gaps 4;  
QY 246 CGAAGCTTCCGCGTGTCCCGCCAGATCGCGAAGCGCTGCGGCGCGCCGCTCA 305  
|||  
Db CGAAGCTTCCGCGTGTCCCGCCAGATCGCGAAGCGCTGCGGCGCGCCGCTCA 9179  
|||  
QY 306 CTTCCCGCAGCGTGGCAGCACCTGCGCTTTGGCCCCCGCGCTCGCGCTCGGCGCTCT 365  
|||  
Db GCTCTTGACCGTGGGACACCGAGCGCTTCGCTCCCGCTCGCGCTCGGCGCTCT 9119  
|||  
QY 366 CCAGGAACCGCTCGCGCAGCGGCTTCGGATATCTCCGAGAGCATCTGCTCGCGCGCGGCT 425  
|||  
Db GCAGGAAGCGCTCCGCGCAGCGGCGGTGGGTACTTCGCCATCAGGGCCGCTACGCGGGCGT 9059  
|||



QY	426	CCAGCCCCAGCTCCTTCGAAACGCAAGTGAACGCGCCCCAGACACTCTCGAGAGACGCGCGGTCT	485
DB	9058	CGAGCCCGAGTTCTTCGAGCGCAACGCGCGCCCGAGCACTCGGGAGAGCTGGCGCGTCT	8599
QY	486	GCTGCTTGGCGGTGATCGCTTCGGGCGCGGACAGCGGTACGCGGTTCCTTCGTGCGCGG	545
DB	8998	GCTCCCGCGCGGTGATCGCTTCGGGCGCGGACAGCGGTACGCGGTTCGCTCGTGTGCGCGG	8939
QY	546	GC CGGGTCAGTGCCTTGACCGCCACTTCCGCGATGTCGCGCGATCGACGAGCGAACCG	605
DB	8938	TACCGGTGAGAGCGGCGACAGCACCGCGCGGACGTGCGCGGATCGACGACGCGACCG	8879
QY	606	GGGACGTGCGGTACAGCGCGCGGACACAGCCGTTCGGAACCGGATGCGCGGCGCCCGACGACA	665
DB	8878	GAGGTCGCGGTACAGGGCGCGAACACGCGCGCGACTGATCCGGGGGCCAGGAGA	8819
QY	666	GC GTTGTGACATGAAGTCTTGGCGCGCAGGAAGTTCAGTCTAGCCGGAATCCGGTGA	725
DB	8818	GGGTGTGGACATGAACGTCTCGGACGCAACGCTCCACGGGACCCCGAGTCCCTGGA	8759
QY	726	CGGCGCGCTCGTTCTCGCGCTCGCGCGGTGATGAAGTCTCGCGTTCGCTTCCTCCA	785
DB	8758	CCGCTCTGTTCTCCCGTGAACGCGGGTGAAGTCTTCGGGTTCGCTTCCTCCA	8699
QY	786	CCGCGAGCATGCAAGCTTCAACAGGTGCGGACGCGCGCTCGCGCGCGCGCGCGCGA	845
DB	8698	CCGCGCATGGAAGCTTCAACAGGTGCGGATGAGCGCGCGCGCGCGCGCGCGCGA	8642
QY	846	AACGCTGCTGCTCGGTCGCTGCGGTCGCTGAGAGGAACGCGCGCGCGCGCGCGCT	905
DB	8641	CCGCTCTGCTCGGGCTCGCTCGGGTCTGCTGCTCAACAGGAACGCGCTGACGCGCC	8582
QY	906	TGAGGGCCCGGTCAGGCGCGGGGTCGGGTCGCTGCTGCGCGCGGACGCTGACGCTTG	965
DB	8581	GCAGCGCCCGGTCGAGAGCGCGGTCGCGCTGAGCGCGCTCGGACGCTGACGCGCTG	8522
QY	966	GGCGGTGAGAGTCAACCGCTTCGCGCGCGCGGCGAGGACTCTGACGGGACCGTCCGG	1025
DB	8521	TCCCCCGCACCGTGAGCGCTTCGGGGCGCGCGGAGGATCGCGAGGTCGCGCATCCG	8462
QY	1026	CCAGCAGGTGGGCGACCTGACGGCGGACCAACACCGGTCAGCGCGGTCAAGAATCACTC	1085
DB	8461	CCAGCGTTCGGCGACCTCAACGCGCCACCGCTCGGTTCGCAACCGGTACAGAAATCACTC	8402
QY	1086	GGGGCTCTCTTCGGGACAGGAGGAGGGGCGCTTCGAACTATAGGGGGAAGGGC	1145
DB	8401	GGGGAACTCT-----8392	8392
QY	1146	AGGATGCCCCGGGGCGGAAACGGGCGATGTTTCGCGCCCCGGGGCGGTGCTTCAGCG	1205
DB	8391	-----TCAGCGG8385	8385
QY	1206	GAGAAACGGGGGCGGACAGCGTGGTGAGGAAGACGAGGCTGCGCTCTGATGCCCGGTG	1265
DB	8384	TGGGCCCGGGGCGGACAGGACCGGTGGAAGACGCTCTCGCGCTCTGATGCGCTGTC	8325
QY	1266	ACCGCGACGCTGCTGACGCCCGCGCGCCCGGGCGGACCGTTCGCTGCTGATTCAGCAC	1325
DB	8324	ACCGGACGCTCAACCGTATCTCGTTCGCGCGCGGGCTGGACCAACCGCCCCCTATTCAGCAC	8265
QY	1326	GGGCTGTAACCTCCGCGTACCGGACGAACCGGGTGGCGGATCGAGCGGACGAGGGGC	1385
DB	8264	GGGCTGCGAACTCGGATACCGGTGGAACCGGGTGGCGGCTTCACCGGAAAGATTCG	8205
QY	1386	GCGGACCGGTCAGAGGCGACCGCGCTGCGCTGCGCTTCGAGCAGCAGCATGCCGCT	1445
DB	8204	GCGGGCCCGCACAGGCGACGCGGCTGAGGGGCGCTTCAGCAGGAGCATGCCGCGG	8145
QY	1446	ACGTGCTGTTGGGGCGCTGGAAGAGGGTTCGGGTGAATGTGTGTCACCGCGAGTGCAC	1505
DB	8144	ACGTGCTGTTGGGGCGCTGGAAGAGGGTTCGGATTCGGGTGTCCACGCGCAGTTCCAG	8085

Qy	1506	GTGTCTCTGCTGGGAACTTCGCGGACAGGACCA	CGTCTCTGTCGGCGGTGCGACCGGCGCGC	1565
Db	8084	ACTCCCTCCCGCCCGTCCCGGAGAGACCA	CGTCTCTGACGCGGCGCGGCGCCGCGCGC	8025
Qy	1566	GCCGCGGGGACACGGGCGCGGTCCGCGGCAC	CGATGTCGCTGGGAGTTCGCGAAGTCCGCGCGC	1625
Db	8024	GAGCCGGGACCGCGCGCGTCTCCGTTAGGGA	TATGCCCTTCGACCGGGACGTTCACGCGC	7965
Qy	1626	ATCCGCGCGGTAGACTTTGAGGACTTGGTGA	AGCGCGTCTGGGCACTCCCGTGGCAGCGAGC	1685
Db	7964	ATCCGCGCGGTAGCGCTTGGGCGTGTGAAC	CGGGTTCGCGGCACACCCCGTTCGCGCGAGC	7905
Qy	1686	CGTTCGGGCGGGGCGACGGGCGCAAGTCCA	CTGTGTCCTGTACCGGGTTCGCGCGCGGAC	1745
Db	7904	CGGTCCCGCGGTGCAACCGGCCAGCGCGCC	CTGCGCGGAGACCGCGGCGCGCAC	7845
Qy	1746	TTTCAGCTTGGGACAGGCGCACTTCCACCTC	TCAGCTCCGCGGACCTCGCGCCGACCGCGAGG	1805
Db	7844	TTTCAGTTCGAGCAGAAACACTTCAGGCCGA	TCTCGTGGGCTTCGCCCCCAGCGCGAGG	7785
Qy	1806	TGCTTCAGGTGGCAGGTGTAGTCCAGCGTGG	CCATCAGGAAGTGTGTAGCGCGGGATC	1865
Db	7784	TGCTCGGATGGCAGACGTAGTCCAGCTCCG	TTCAGGAGAGTGTGTAGCGCGGGATC	7725
Qy	1866	CCGTAGCCGGCGTGGGAAGCAGAGCATCGCC	CGCTGACGAGGTCTCGCGGATCAGGAGC	1925
Db	7724	CCGTACCCGGCGTGGGAAGCGAGCATGGCCG	CTGCGCATGGCCTCCGCGACCGAGCAGC	7665
Qy	1926	GGATCTGTCCGTCCCGCTGGACCGGTGCGA	AGAACGGGTGTGTGGGCGACGACGGCG	1985
Db	7664	GGATCTGTGCAAGTCTGCAACCCAGGAGCG	AAAGGATGTGTGTCGCGAAGACCGCT	7605
Qy	1986	GTCAACCGAGAACCGGTCAACCGCCCTTTCG	GTATCCAGTGCACCGGGAAACCGCTCTGCACC	2045
Db	7604	TCACACCGGNAACCGGTTCGCGCCCGAGCG	GTACCCAGTTCGCGGAAAGCGCTCTCCGGC	7545
Qy	2046	CTGGTCCGGTGGACAAGCGCCATCGGAACG	CGCAATGCGGTTTGTTCGATCGAGTTGGCA	2105
Db	7544	CTGGTCCGGTGCACCATCTCGATCCCGAC	CGGATGCACACCTTCGGCTCCGGTTCCATTA	7485
Qy	2106	TCGAGCGCAGATTGATCAAACTACTGCTTCG	GGCATGGGTCCCGCCAGGATCATGT	2165
Db	7484	GTCCGACCCGGATCGATCAAGACTGTGCTTC	CGGCGATAACT-CCCCCGAGGGGTCTATGT	7426
Qy	2166	GATGCGGAGCTGTC-----TGATGCGGAA	CGTTAAGATACAGACTGAGCGGTTTT	2217
Db	7425	AGTGGCGAGCGTTCTCGATCGCGGTACTGT	GATAGACAGATACAGATATATCGTTTT	7366
Qy	2218	TTTTT 2221		
Db	7365	TTTTT 7362		
RESULT 4				
AB088224/c				
LOCUS 210614 bp DNA linear BCT 11-JUN-2003				
DEFINITION Streptomyces rochei plasmid pSLA2-L DNA, complete sequence.				
ACCESSION AB088224				
VERSION AB088224.1 GI:30698345				
KEYWORDS				
SOURCE Streptomyces rochei				
ORGANISM Streptomyces rochei				
1 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
Streptomycineae; Streptomycetaceae; Streptomyces.				
2				
REFERENCE				
1 Kinashi,H., Fujii,S., Hatan,A., Kurokawa,T. and Shinkawa,H.				
AUTHORS Physical mapping of the linear plasmid pSLA2-L and localization of				
TITLE the eryAI and actI homologs				
JOURNAL Identification of two polyketide synthase gene clusters on the				
AUTHORS Biosci. Biotech. Biochem. 62, 1892-1897 (1998)				
2				
Suwa,M., Sugino,H., Sasaoka,A., Mori,E., Fujii,S., Shinkawa,H.,				
Nimi,O. and Kinashi,H.				
TITLE Identification of two polyketide synthase gene clusters on the				

JOURNAL PUBMED REFERENCE AUTHORS TITLE	linear plasmid pSLA2-L in Streptomyces rochei	CDS	<p>ARCFTLTSEAIDRGICAPVQVVCVDTITDQLQAQQLGVGRSDEVRGARLAALQTA LLKASSENFRRLTVFHHVKEARFAAGLPDVAKRHAAGPGLYPRTIWAWLGGSH KPHRRSEVLGEFTSGIATDGTVEKGLSVKVLGEGVDRECDVYVWADVRCGMPDL VQAVGRALRIPOCGQKVASLVPLDLPGETADNMLTSRPFYNGIALLFRCQAPLITGHG EGFRAARFPFGFVRPVLGVGRQAGSAGASPSRSITPSMSLSISMRAP"</p> <p>complement (2315. .3595) /note="ORF2 (426 aa) similar to AB004736-10 Pseudomonas aeruginosa hypothetical protein (442 aa)" /codon_start=1 /transl_table=11 /product="putative secreted protein" /protein_id="BAC76460.1" /db_xref="GI:30698347"</p> <p>/translation="MRKPRALLGIPGLLAVLAPAPAAHAGADTVGVWAAPTTAA PAADSTVEYDQTLRQTVHLSVAGDSLRVRETFGTSPLTIGVHARPAAGPATV DPGTVRVREGPSATLAEPTQKSDPVALPTTAGDGLVLSLVPRTGSGTSHSAA YQNFVAAGDVTGAPDLTPVSTATSMHFLSGVSDRRAGTADSSVLTGOSITDGEHT TLDAKRRWPDLLAERLRDGLAGTVGNAGIGNRLLRDPPDPGSAASFYAFGE SALKRDRDVLGQPGARAVTVLGVNDLQGPAAAPASDEVTAELIAGYRQLIERAH EHGLKLYGATITPAGDTICGYFTPRRAVRQVNDWRTSGAFDTLDFDAVLRDPAR PDHLLPAYDGGDLHPNDAGMAAMARAFPLDSLRL"</p> <p>complement (3659. .4645) /note="ORF3 (328 aa) similar to Y00459-2 Streptomyces griseus regulatory protein, StrR (350 aa)" /codon_start=1 /transl_table=11 /product="StrR-like regulatory protein" /protein_id="BAC76461.1" /db_xref="GI:30698348"</p> <p>/translation="MTAALAGNDQETRSIPISLVPGYSPRLCGEAEHVARLADTD EALPPLIVESTLRVLDGHEVLAACKGHTIEVRLFDGAEEAFLLAVRSNTHGL PLSRDQRRRAAQILQWPHLSDBAVAGIAGIKTVAAIRPLAAGTTPQARGRD GRRLPDQTRRRKAAELAQPAQSVREVARHAGISPATASDVRRILASGRSPYPER NTTCARPAAGTTSRATPGGAPEPVSVPIRPPEPSPVLRLDPSLRHSGRRL LRLQCAVERTALLAQAQVTPHCTDLVAELAREYADLWAEFAREVATDG"</p> <p>complement (4922. .5812) /note="ORF4 (296 aa), lankacidin biosynthesis protein similar to AE004625-4 Pseudomonas aeruginosa pyrroloquinoline quinone biosynthesis protein B (304 aa)" /codon_start=1 /transl_table=11 /product="pyrroloquinoline quinone biosynthesis protein B" /protein_id="BAC76462.1" /db_xref="GI:30698349"</p> <p>/translation="MILLGTAAAGGPPQWNCACALCARGRGELPARSQECVAVSGDG RDMWLLNASPDRLTQLLAAPALTPEGPRDTPVRGVLLTDAEVHAGLVIRATGL TVYAAPVPRGDSIAELPVRGLLDRYAPDWDRDAPGGAAGGLTVTAHPVGTAKPK YAHAPDPDAPWVCAYRIEDPATGGALVYAPCLATWPDGFDLLASATCALLDGTFPSA GELGTATSSAGAGOSLMGHLFVAGPGGSLAALARRHGLRRIYTHLNTNPLLDPSAA HAAVREAGVEVLPDGSSELVL"</p> <p>complement (5815. .6909) /note="ORF5 (364 aa), lankacidin biosynthesis protein similar to AJ277117-5 Gluconobacter oxydans putative pqg protein (359 aa)" /codon_start=1 /transl_table=11 /product="pyrroloquinoline quinone biosynthesis protein E" /protein_id="BAC76463.1" /db_xref="GI:30698350"</p> <p>/translation="MADPAVGCAPAGMLIETHRCPLHCPCYCNPLFVRRAEALTCEQ WTDLTQARELGVWMPDSGCEPLDPLVGHARRLGAVNLVTSVGVLTAEARAH DIARRGVHVSLSQDADPAQQAIALAGAVHTAKLEAAVATAAGLPTVNLVLRGN IDRTGRMVDLADIGADRIELANTQYVWGLNRNALMPTAAQLAAAREAVRHATRY AGDEPLVVAADYDDPKCMGWSQTLTVTAGDVLPCPFAATITLIPVENALRR PLSEIWTASRSFNAYRGVGNRREPCRTCTPCHADHGGCRCAQFQLTGDAADTDPACGL SPHSLVDAALAEVTDGVPVAFVPRGPVA"</p> <p>complement (6902. .7174) /note="ORF6 (90 aa), lankacidin biosynthesis protein similar to AL603642-200 Sinorhizobium meliloti putative pyrroloquinoline quinone biosynthesis protein D (98 aa)" /codon_start=1</p>
	Gene 246 (1-2), 123-131 (2000)		
	10767533		
JOURNAL PUBMED REFERENCE AUTHORS TITLE	Hiratsu, K., Mochizuki, S. and Kinashi, H. Cloning and analysis of the replication origin and the telomeres of the large linear plasmid pSLA2-L in Streptomyces rochei	CDS	<p>complement (5815. .6909) /note="ORF5 (364 aa), lankacidin biosynthesis protein similar to AJ277117-5 Gluconobacter oxydans putative pqg protein (359 aa)" /codon_start=1 /transl_table=11 /product="pyrroloquinoline quinone biosynthesis protein E" /protein_id="BAC76463.1" /db_xref="GI:30698350"</p> <p>/translation="MADPAVGCAPAGMLIETHRCPLHCPCYCNPLFVRRAEALTCEQ WTDLTQARELGVWMPDSGCEPLDPLVGHARRLGAVNLVTSVGVLTAEARAH DIARRGVHVSLSQDADPAQQAIALAGAVHTAKLEAAVATAAGLPTVNLVLRGN IDRTGRMVDLADIGADRIELANTQYVWGLNRNALMPTAAQLAAAREAVRHATRY AGDEPLVVAADYDDPKCMGWSQTLTVTAGDVLPCPFAATITLIPVENALRR PLSEIWTASRSFNAYRGVGNRREPCRTCTPCHADHGGCRCAQFQLTGDAADTDPACGL SPHSLVDAALAEVTDGVPVAFVPRGPVA"</p> <p>complement (6902. .7174) /note="ORF6 (90 aa), lankacidin biosynthesis protein similar to AL603642-200 Sinorhizobium meliloti putative pyrroloquinoline quinone biosynthesis protein D (98 aa)" /codon_start=1</p>
	Mol. Gen. Genet. 263 (6), 1015-1021 (2000)		
	10954087		
JOURNAL PUBMED REFERENCE AUTHORS TITLE	Mochizuki, S., Hiratsu, K., Suwa, M., Ishii, T., Sugino, F., Yamada, K. and Kinashi, H. The large linear plasmid pSLA2-L of Streptomyces rochei has an unusually condensed gene organization for secondary metabolism	CDS	<p>complement (5815. .6909) /note="ORF5 (364 aa), lankacidin biosynthesis protein similar to AJ277117-5 Gluconobacter oxydans putative pqg protein (359 aa)" /codon_start=1 /transl_table=11 /product="pyrroloquinoline quinone biosynthesis protein E" /protein_id="BAC76463.1" /db_xref="GI:30698350"</p> <p>/translation="MADPAVGCAPAGMLIETHRCPLHCPCYCNPLFVRRAEALTCEQ WTDLTQARELGVWMPDSGCEPLDPLVGHARRLGAVNLVTSVGVLTAEARAH DIARRGVHVSLSQDADPAQQAIALAGAVHTAKLEAAVATAAGLPTVNLVLRGN IDRTGRMVDLADIGADRIELANTQYVWGLNRNALMPTAAQLAAAREAVRHATRY AGDEPLVVAADYDDPKCMGWSQTLTVTAGDVLPCPFAATITLIPVENALRR PLSEIWTASRSFNAYRGVGNRREPCRTCTPCHADHGGCRCAQFQLTGDAADTDPACGL SPHSLVDAALAEVTDGVPVAFVPRGPVA"</p> <p>complement (6902. .7174) /note="ORF6 (90 aa), lankacidin biosynthesis protein similar to AL603642-200 Sinorhizobium meliloti putative pyrroloquinoline quinone biosynthesis protein D (98 aa)" /codon_start=1</p>
	Mol. Microbiol. 48 (6), 1501-1510 (2003)		
	12791134		
JOURNAL PUBMED REFERENCE AUTHORS TITLE	5 (bases 1 to 210614) Mochizuki, S., Hiratsu, K. and Kinashi, H. Direct Submission	CDS	<p>complement (5815. .6909) /note="ORF5 (364 aa), lankacidin biosynthesis protein similar to AJ277117-5 Gluconobacter oxydans putative pqg protein (359 aa)" /codon_start=1 /transl_table=11 /product="pyrroloquinoline quinone biosynthesis protein E" /protein_id="BAC76463.1" /db_xref="GI:30698350"</p> <p>/translation="MADPAVGCAPAGMLIETHRCPLHCPCYCNPLFVRRAEALTCEQ WTDLTQARELGVWMPDSGCEPLDPLVGHARRLGAVNLVTSVGVLTAEARAH DIARRGVHVSLSQDADPAQQAIALAGAVHTAKLEAAVATAAGLPTVNLVLRGN IDRTGRMVDLADIGADRIELANTQYVWGLNRNALMPTAAQLAAAREAVRHATRY AGDEPLVVAADYDDPKCMGWSQTLTVTAGDVLPCPFAATITLIPVENALRR PLSEIWTASRSFNAYRGVGNRREPCRTCTPCHADHGGCRCAQFQLTGDAADTDPACGL SPHSLVDAALAEVTDGVPVAFVPRGPVA"</p> <p>complement (6902. .7174) /note="ORF6 (90 aa), lankacidin biosynthesis protein similar to AL603642-200 Sinorhizobium meliloti putative pyrroloquinoline quinone biosynthesis protein D (98 aa)" /codon_start=1</p>
	Submitted (15-JUN-2002) Haruyasu Kinashi, Hiroshima University, Department of Molecular Biotechnology, Graduate School of Advanced Sciences of Matter; 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8530, Japan (E-mail: kinashi@hiroshima-u.ac.jp, Tel: 81-824-24-7869, Fax: 81-824-24-7869)		
	739-8530, Japan (E-mail: kinashi@hiroshima-u.ac.jp, Tel: 81-824-24-7869, Fax: 81-824-24-7869)		
JOURNAL PUBMED REFERENCE AUTHORS TITLE	The nucleotide sequence has been determined by using restriction fragments and nested deletion fragments of the ordered cosmid library of pSLA2-L. pSLA2-L is 210,614 bp long with G+C content of 72.8 % and carries 143 ORFs.	CDS	<p>complement (5815. .6909) /note="ORF5 (364 aa), lankacidin biosynthesis protein similar to AJ277117-5 Gluconobacter oxydans putative pqg protein (359 aa)" /codon_start=1 /transl_table=11 /product="pyrroloquinoline quinone biosynthesis protein E" /protein_id="BAC76463.1" /db_xref="GI:30698350"</p> <p>/translation="MADPAVGCAPAGMLIETHRCPLHCPCYCNPLFVRRAEALTCEQ WTDLTQARELGVWMPDSGCEPLDPLVGHARRLGAVNLVTSVGVLTAEARAH DIARRGVHVSLSQDADPAQQAIALAGAVHTAKLEAAVATAAGLPTVNLVLRGN IDRTGRMVDLADIGADRIELANTQYVWGLNRNALMPTAAQLAAAREAVRHATRY AGDEPLVVAADYDDPKCMGWSQTLTVTAGDVLPCPFAATITLIPVENALRR PLSEIWTASRSFNAYRGVGNRREPCRTCTPCHADHGGCRCAQFQLTGDAADTDPACGL SPHSLVDAALAEVTDGVPVAFVPRGPVA"</p> <p>complement (6902. .7174) /note="ORF6 (90 aa), lankacidin biosynthesis protein similar to AL603642-200 Sinorhizobium meliloti putative pyrroloquinoline quinone biosynthesis protein D (98 aa)" /codon_start=1</p>
	Gene prediction was based on the unique codon usage in Streptomyces (Bibb et al., Gene 30:157-66 (1984)) using the Frameplot program of Ishikawa and Hotta (FEMS Microbiol Lett 174: 251-253 (1999)) as implemented at <a href="http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl">http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl</a> . Where possible we chose an initiation codon (atg, gtg, ctg or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13 bp before the initiation codon). If this could not be identified we chose the most upstream initiation codon.		
	Three-fourths of the pSLA2-L DNA are occupied by biosynthetic genes (three PKS gene clusters for lankacidin, lankamycin and an unknown type II polyketide, a carotenoid biosynthetic gene cluster, many regulatory genes and others). The range of each biosynthetic gene cluster has been deduced by comparison with similar gene clusters in most cases and may be revised in future.		
FEATURES source	Location/Qualifiers	CDS	<p>complement (5815. .6909) /note="ORF5 (364 aa), lankacidin biosynthesis protein similar to AJ277117-5 Gluconobacter oxydans putative pqg protein (359 aa)" /codon_start=1 /transl_table=11 /product="pyrroloquinoline quinone biosynthesis protein E" /protein_id="BAC76463.1" /db_xref="GI:30698350"</p> <p>/translation="MADPAVGCAPAGMLIETHRCPLHCPCYCNPLFVRRAEALTCEQ WTDLTQARELGVWMPDSGCEPLDPLVGHARRLGAVNLVTSVGVLTAEARAH DIARRGVHVSLSQDADPAQQAIALAGAVHTAKLEAAVATAAGLPTVNLVLRGN IDRTGRMVDLADIGADRIELANTQYVWGLNRNALMPTAAQLAAAREAVRHATRY AGDEPLVVAADYDDPKCMGWSQTLTVTAGDVLPCPFAATITLIPVENALRR PLSEIWTASRSFNAYRGVGNRREPCRTCTPCHADHGGCRCAQFQLTGDAADTDPACGL SPHSLVDAALAEVTDGVPVAFVPRGPVA"</p> <p>complement (6902. .7174) /note="ORF6 (90 aa), lankacidin biosynthesis protein similar to AL603642-200 Sinorhizobium meliloti putative pyrroloquinoline quinone biosynthesis protein D (98 aa)" /codon_start=1</p>
	1. .210614		
	/organism="Streptomyces rochei" /mol_type="genomic DNA" /strain="7434AN4" /db_xref="taxon:1928" /plasmid="pSLA2-L" /note="linear plasmid"		
LTR	1. .1992	LTR	<p>complement (5815. .6909) /note="ORF5 (364 aa), lankacidin biosynthesis protein similar to AJ277117-5 Gluconobacter oxydans putative pqg protein (359 aa)" /codon_start=1 /transl_table=11 /product="pyrroloquinoline quinone biosynthesis protein E" /protein_id="BAC76463.1" /db_xref="GI:30698350"</p> <p>/translation="MADPAVGCAPAGMLIETHRCPLHCPCYCNPLFVRRAEALTCEQ WTDLTQARELGVWMPDSGCEPLDPLVGHARRLGAVNLVTSVGVLTAEARAH DIARRGVHVSLSQDADPAQQAIALAGAVHTAKLEAAVATAAGLPTVNLVLRGN IDRTGRMVDLADIGADRIELANTQYVWGLNRNALMPTAAQLAAAREAVRHATRY AGDEPLVVAADYDDPKCMGWSQTLTVTAGDVLPCPFAATITLIPVENALRR PLSEIWTASRSFNAYRGVGNRREPCRTCTPCHADHGGCRCAQFQLTGDAADTDPACGL SPHSLVDAALAEVTDGVPVAFVPRGPVA"</p> <p>complement (6902. .7174) /note="ORF6 (90 aa), lankacidin biosynthesis protein similar to AL603642-200 Sinorhizobium meliloti putative pyrroloquinoline quinone biosynthesis protein D (98 aa)" /codon_start=1</p>
	/note="left terminal inverted repeat, TIR-L; shows 99.4 % (1981/1992) sequence identity to TIR-R (complement (208623. .210614))"		
	683. .2188		
CDS	/note="N-terminal sequence is almost identical (435/437) with that of ORF143 at the right end of pSLA2-L until the inner end of TIR-L ORF1 (501 aa) similar to AL590463 Streptomyces coelicolor putative helicase, SCPI.136 (879 aa); homology is seen until the inner end of TIR-L"	CDS	<p>complement (5815. .6909) /note="ORF5 (364 aa), lankacidin biosynthesis protein similar to AJ277117-5 Gluconobacter oxydans putative pqg protein (359 aa)" /codon_start=1 /transl_table=11 /product="pyrroloquinoline quinone biosynthesis protein E" /protein_id="BAC76463.1" /db_xref="GI:30698350"</p> <p>/translation="MADPAVGCAPAGMLIETHRCPLHCPCYCNPLFVRRAEALTCEQ WTDLTQARELGVWMPDSGCEPLDPLVGHARRLGAVNLVTSVGVLTAEARAH DIARRGVHVSLSQDADPAQQAIALAGAVHTAKLEAAVATAAGLPTVNLVLRGN IDRTGRMVDLADIGADRIELANTQYVWGLNRNALMPTAAQLAAAREAVRHATRY AGDEPLVVAADYDDPKCMGWSQTLTVTAGDVLPCPFAATITLIPVENALRR PLSEIWTASRSFNAYRGVGNRREPCRTCTPCHADHGGCRCAQFQLTGDAADTDPACGL SPHSLVDAALAEVTDGVPVAFVPRGPVA"</p> <p>complement (6902. .7174) /note="ORF6 (90 aa), lankacidin biosynthesis protein similar to AL603642-200 Sinorhizobium meliloti putative pyrroloquinoline quinone biosynthesis protein D (98 aa)" /codon_start=1</p>
	/codon_start=1		
	/product="putative helicase" /protein_id="BAC76459.1" /db_xref="GI:30698346"		
JOURNAL PUBMED REFERENCE AUTHORS TITLE	/translation="MSTTSRTDQREAAQOAEVAVVRALELPVRLAPERGLRTQVIM ATGSKTVAARSASAKRAGRLVPLVPSLDLITQTEAAWREAGRTGPMIGVSSLRGED VAFPNNTVEELVDVVRPDKVTVFATVASLGLTLEAHRGGLPGWDLIVDEAHT SGRLGKPVAVHDNTRIFSLRLRYNTATFRLWQLEDDEAGAGELVASMEDDDPDGLFG	CDS	<p>complement (5815. .6909) /note="ORF5 (364 aa), lankacidin biosynthesis protein similar to AJ277117-5 Gluconobacter oxydans putative pqg protein (359 aa)" /codon_start=1 /transl_table=11 /product="pyrroloquinoline quinone biosynthesis protein E" /protein_id="BAC76463.1" /db_xref="GI:30698350"</p> <p>/translation="MADPAVGCAPAGMLIETHRCPLHCPCYCNPLFVRRAEALTCEQ WTDLTQARELGVWMPDSGCEPLDPLVGHARRLGAVNLVTSVGVLTAEARAH DIARRGVHVSLSQDADPAQQAIALAGAVHTAKLEAAVATAAGLPTVNLVLRGN IDRTGRMVDLADIGADRIELANTQYVWGLNRNALMPTAAQLAAAREAVRHATRY AGDEPLVVAADYDDPKCMGWSQTLTVTAGDVLPCPFAATITLIPVENALRR PLSEIWTASRSFNAYRGVGNRREPCRTCTPCHADHGGCRCAQFQLTGDAADTDPACGL SPHSLVDAALAEVTDGVPVAFVPRGPVA"</p> <p>complement (6902. .7174) /note="ORF6 (90 aa), lankacidin biosynthesis protein similar to AL603642-200 Sinorhizobium meliloti putative pyrroloquinoline quinone biosynthesis protein D (98 aa)" /codon_start=1</p>
	Gene 246 (1-2), 123-131 (2000)		
	10767533		



```
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein D"
/protein_id="BAC76464.1"
/db_xref="GI:30698351"
/translation="MTGLPEPVPRLRQVRLTRDPARGELALLPERVVVLDNTAAAV
LAHCDGTTSLAGIVRLAEVGVSAEDVRELLLLAQRVVLDLG"
complement(7171..7899)
/note="ORF7 (242 aa), lankacidin biosynthesis protein
similar to AE004625-5 Pseudomonas aeruginosa
pyrroloquinoline quinone biosynthesis protein C (250 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein C"
/protein_id="BAC76465.1"
/db_xref="GI:30698352"
/translation="MSMSTVREVAAPWSEAEPRORLHALESSYMDRHPFRRMHGELL
DEGELRLWAANPWYORCLPKDAAIVANCPLEVRROWLSRIYVHDGADACAGAEK
WURLABVGLRDEVDHRLVLAGTRFAVDAYDFARRPWLAAASGLITELFSPLL
AHLRLREHYFWIAEBGFYFARIEVVVPGRSLDLDLVANHAVRSQECVRLA
FKRVLNAVLDLSLDYHTNGATRS"
complement(7926..8030)
/note="ORF8 (34 aa), lankacidin biosynthesis protein
similar to AL603642-197 Sinorhizobium meliloti putative
pyrroloquinoline quinone biosynthesis protein A (31 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein A"
/protein_id="BAC76466.1"
/db_xref="GI:30698353"
/translation="MRTSGKELPAKKAMHRPFDVITDGMETVAFPSR"
8420..10057
/note="ORF9 (545 aa), possible lankacidin resistance
protein
similar to AL133424-19 Streptomyces coelicolor putative
ABC transporter ATP-binding protein (544 aa)"
/codon_start=1
```

Query Match 13.0%; Score 563; DB 1; Length 210614;

Best Local Similarity 62.8%; Pred. No. 1.8e-40;

Matches 981; Conservative 0; Mismatches 535; Indels 47; Gaps 5;

QY	2260	CATGGCAAGCAGGACCGGCGATCCGACGGGCGAGAGATCCTGGAGCGCGCGCGCA	2319
DB	143125	CGTGGCACAGCAGGAACGGGCGATCCGACCGCGCGCGGTCTCGAAGCGCAGCCAC	143066
QY	2320	GGTCTTCGAGACGAGGCTACCAAGTCGCAGATCAGGAGATCTCAAGTGGCGCG	2379
DB	143065	GGTATTCGCGGACGAGGCTAGCCCGCGACCGGTGGCGGACATCTCAAGTGGCGCG	143006
QY	2380	GGTGACCAAGGAGCCCTCTACTTCCACTTCCAGTCCAAAGGAAGAACTGGCGTGGGCGT	2439
DB	143005	CCTGACCAAGGAGCGCTCTACTTCCACTTCCAGTCCAAAGGAAGCGCTTGGCGGCGCAT	142946
QY	2440	CTTTCGACCGCCAGGAACCAACAGCGCGTTCGAGAGCAACCCCTCCGCGTGCAGAACT	2499
DB	142945	CCTGGAGGCGCCAGGTGCGCAGCAGCTCGTTCGCGAGCA---GCTCAAGCTCCAGGAATG	142889
QY	2500	CATCGACATGGGCGATGTTGTTCTGTCTCAGCTTGGCGAGCAAGAGTCTGGCGCGGCGCG	2559
DB	142888	GGTGGACGGGGGATGACGCTGGCGACCAAGCTGCGCGGGAGCCCGCTGTGCGCGCGG	142829
QY	2560	CGTGGCGCTCTCCATGACGACGAGCGGACGCGTCTCGATCGCGGAGGACCCCTTCGCTCG	2619
DB	142828	CGCCCGGCTCTTCGGCGGAGCAACCGGAGCGAGC---AGCAGGAGCGGCTTCCCGAC	142772
QY	2620	CTGGCAGCAGACACTCTCTGAAGCTGCTGAACAGGCGCAAGGAGAACGGTGAAGTGTGCC	2679
DB	142771	CTGGATCGGCTTCTCGGCGCTCGCTCTCTTGAGCAGGCGAAGCGGAAACGGGAGGTCTCTGG	142712
QY	2680	CCATGTGGTCAACACCGATCTGGCGGATCTCTACGTGGGACGTTTCGCGGGATACAGT	2739
DB	142711	CCACATCAGCGCGCGGACCGCGGAATGCGTCTCGGCTCTCTCCACGGCATCAGCT	142652

QY	2740	CGTGTCCAGACGGTACGGACTACAGACCTCGAAACACCGCTACGCGCTGCTGCAGAA	2799
DB	142651	CCTCTCCAGTTGCAGACGAATCGGCGGACATCAGCAGCGTGCTTCGCGCTGTTCGG	142592
QY	2800	GCACATCTGCGCGGCATCGCGGTTCCCTCCGTGCTGGCGCGCTCGATCTCTCCGAGGA	2859
DB	142591	GCACGTCTCTCCGCGCGTGGCGGTCCCTCGGTGCTGGTGGCGGTGGACACCGCCCGGA	142532
QY	2860	GGCGGAGACAGCTTCGCGGCGGAATGGCAACCG-----	2897
DB	142531	CGGGGGGCGCGTCTGTCGCGAGCTGAGGCGATGGCGGCGCAGCCGAGCGGCTGCG	142472
QY	2898	-----GGAAGAATGACCGCGGAGCGCCCGCACCGGATACC--GACCGCGC	2942
DB	142471	CTCGTGTGAGCTGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	142412
QY	2943	GTGCCCGAGCG	3002
DB	142411	CGGCG	142352
QY	3003	TGCGTACCGAAGCGTGGCGGCT---CAGAGAAATCGTTCGCGCTGTGGCATCGAGTACT	3058
DB	142351	CTTGCCCAACCATCTTTTGGGAGTCTTACAGCGCCCGCGCGCGCGCGCGCGCGCGCG	142292
QY	3059	GGCGGCTGATCACCGTGAAGTCTGGAGGCGAGAAAGCGCACCGCTGGCGGATGCTGT	3118
DB	142291	GACCGGTGATCACCGGAGTCTGCGAGGCGAGAAAGCGCGAGATGTCGGCGATGTCG	142232
QY	3119	CGGGTCTCGGATGCGGTTGAACAGCGGATTCGGCGCGCAGTCTCGCGCGCGCGCTCGGGG	3178
DB	142231	CGGTCTCTCCATGCGCGGAGAGACCGAGGCGCGCGTGGCGCGCGCTCGCGCGCGCTCGGGTG	142172
QY	3179	TCGCGCGCGCGCGTGTCTGATGCTCTTCCAGAAACCGCGCGCGCGCGCGCGCGCGCG	3238
DB	142171	TGGCCCGAGGTCGCGGCTTCACTGCTTCAAGAAATCGCGCGCGCGCGCGCGCGCGCG	142112
QY	3239	TGATCCCGGTTCCCGCGAGTTCCTGGCGAGGCGAGCGTGTGTCCACCGCACCTT	3298
DB	142111	TGATCCCG	142052
QY	3299	TGGTCACTCGGATCCGAGTCTCGGGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG	3358
DB	142051	TGCGCATGCGATGAGGATGAGTCTGGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG	141992
QY	3359	TGATGATCCGCGCGCGTGGCGAGTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	3418
DB	141991	TGACGATCCGTCG	141932
QY	3419	CCGCGAGTTCAGGCGGACAGTTCGTAAGACCTCTCTCGGTGCTTCCGTGATCCGTC	3478
DB	141931	CCTTGGCGTGCAGGCGCATCAGCTCGTAAGTCCGCGCGCGCGCGCGCGCGCGCGCG	141872
QY	3479	CGGAGCGGTCAGCGCGCGGTTGTTTCCAGGATGTGGAATCTCGGCGCGCGCGCGCG	3538
DB	141871	CCGCGTGTGTTGTTCCCGGTTTTCACCGAGATGTGAGGCGCGCGCGCGCGCGCGCG	141812
QY	3539	CGCCCATCCCGCGCGTGCAGACCGCGGTAGCGCGCGCGCGCGCGCGCGCGCGCGCG	3598
DB	141811	CCCTCAGCGCGGTCGATCGCTGTACAGGTTCGCGCGCGCGCGCGCGCGCGCGCGCG	141752
QY	3599	CGGCGCGGATGCGCAACCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	3658
DB	141751	GGGCGTGCACCATGACG	141692
QY	3659	CGCGCGCTCGTGTGCGGTAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	3718
DB	141691	CGCGCGTCTGTCGCGCGCGGTAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	141632
QY	3719	GGGCGGATACCGGTTCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	3778
DB	141631	GGGCGGATGCGCGCGCGGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	141572
QY	3779	GCG 3781	

Db	141571	CGC 141569			
RESULT 5					
STWAFSAA/c					
LOCUS					
DEFINITION	STWAFSAA	1203 bp	DNA	linear BCT 26-APR-1993	
ACCESSION	S.griseus afsA	gene encoding a possible A-factor biosynthesis protein.			
VERSION	M24250.1	GI:153148			
KEYWORDS	A-factor biosynthesis.				
SOURCE	Streptomyces griseus				
ORGANISM	Streptomyces griseus				
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
AUTHORS	1 (bases 1 to 1203)				
TITLE	Horinouchi, S., Suzuki, H., Nishiyama, M. and Beppu, T.				
	Nucleotide sequence and transcriptional analysis of the				
	Streptomyces griseus gene (afsA) responsible for A-factor				
	biosynthesis				
JOURNAL	J. Bacteriol. 171 (2), 1206-1210 (1989)				
COMMENT	2492509				
FEATURES	Original source text: S.griseus DNA.				
	Location/Qualifiers				
	1..1203				
	/organism="Streptomyces griseus"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:1911"				
	204..1109				
	/note="afsA protein"				
	/codon_start=1				
	/transl_table=11				
	/protein_id="AAA26693.1"				
	/db_xref="GI:153149"				
CDS	translation="MDAAEVVHPVBIEMVHTRPEDAFPRNWRLGRDPAVEAVLP				
	HDHPFPAPVGDLDHDLVAEMQAMLAFAHAGYIPLGYHELLTLDYVCHPEHLG				
	VGBPTEIGLEVFCSLLKWRAGLPAQGRVGHVGRDLAATGVNATFSPFKAYRM				
	RGVDFVEGISLPTAPVPASFAGRARVEDVVLSTGRGVWELRVDTKHPFLFORPND				
	HVPGMLEAARQAACLVAGPAGIVPVEARTFRHYSBFGSPCWIGAVVQGADEDTV				
	TVRVTGHQDGETVFSVLSPRAHG"				
ORIGIN					
Query Match	12.1%;	Score 527.4;	DB 1;	Length 1203;	
Best Local Similarity	70.9%;	Pred. No. 1.3e-36;			
Matches 731;	Conservative	0;	Mismatches 291;	Indels 9;	Gaps 2;
QY	1199	TCAGCGGAGAACCGGGCGGACAGCGTGTGTGAGGAAGACGAGGCTGCCGTCTCTGATG	1258		
Db	1109	TCAGCGGTGGCCCGGGGGCGGACAGGACCGTGGAGAGACCGTCTCGCGTCTCTGATG	1050		
QY	1259	CCCGGTGACCCGACCGGTGTGTCAGCCCCCGCGCCCGGCGGACCGTGGCTGTGATG	1318		
Db	1049	GCCTGTACCCGGACGGTCACCGTATCTCTGTCGCCCGCCCGCTGACACCGCCCTAT	990		
QY	1319	CCAGCACGGCTGTGAACTCCGCTACCGAGCAACCGGTGCCGCGCATCGACGGCAC	1378		
Db	989	CCAGCACGGGTGCCGAACCTCGAGTACCGGTGGAACCGGGTCCGCGCTCCACCGGAAC	930		
QY	1379	GAAGGGCCCGGACCGGTTCAGGACGACCGCCCTCGGTGCCGCTTCGAGCAGCAGCAT	1438		
Db	929	GATTCGGCGGGCCCGCCACACAGAGACGCGGCTTCAGCGCGCGCTTCAGCAGAGGAT	870		
QY	1439	GCCCGGTACGTGGTCTGTTGGGCGCTGGAAGAGGTTCGGGTGACTGGTGTCCACCCGAG	1498		
Db	869	GCCCGGAGCGTGTGTTGGGCGCTGGAAGAGGTTCGATGTCGGGTGTCACGCGCAG	810		
QY	1499	TCGCCACGTGTCTCTGTGGAACTCGCCGACAGGACCGTCTCTGTTCGGGTTCGACC	1558		
Db	809	TTCCAGAGATCTCCTCCCGCCCGTCCCGAGAGACCACTCTCTCGACGCGGGCGGCC	750		
QY	1559	GCGCGCGCGCGGCGACGCGGCGCGTCCCGGCGACCGATGGGTGGGAGTTCGCGAAGTC	1618		
Db	141571	CGC 141569			
QY	1619	GC CGCGCATCCGCGGTAGACTTTCAGGACTGGTGAAGGGGTCTGTCAGACTCCCGTGGC	1678		
Db	689	ACC CGCATCCGCGGTAGGCTTTGGGCGTGTCTGAACCGGGTCCCGCCACCCCCGTGCG	630		
QY	1679	AGC GAGCGCTCCGCGCGGCGACCGGCCAGTCCACTGTCTCCCTGTACGGGCTGCCCGCC	1738		
Db	629	GGC GAGCGGTGCGCCCGGTGCACCGCCAGCCGCGCCCTTCGCGCGGAGCCCGCG	570		
QY	1739	GC GAACTTCAGCTGGGAAACAGGCCACTTCCACTCCAGTTCGCGACCTCGCCGACAC	1798		
Db	569	CCGCCACTTCAGGTCCGAGCAGAACACCTCCAGGCCGATCTCGGTGGGCTCGCCGCCAC	510		
QY	1799	GCC GAGGTTCGAGGTGGCAGGTGTAGTCCAGGTGGCCATCAGGAAGTGTAGCCGAC	1858		
Db	509	GCC GAGGTTCGAGGTGGCAGGTGTAGTCCAGGTTCGAGGTTCGAGGTTCGAGGTTCGAG	450		
QY	1859	CGC CACCGGTAGCGCGGTGGAAAGACGAGCATCGCCGCTTCGACGCGAGGTCTCGGCGAT	1918		
Db	449	CGGATCCCGTACCGCGGTGGAAAGGAGCATGCGCGCTTCGCGCATGGCTTCGCGAC	390		
QY	1919	CAG CAGCGGATCGTTCGTCCTCCCGTGGACCGGTGCGAAGAACCGGTGTCGTGGGGGAG	1978		
Db	389	CAG CAGCGGATCGTTCGTCCTCACCCACCGGAGCGAAGAGGATGGTTCGTCGGAAG	330		
QY	1979	GAC GCGGTGACCGAGAACCGGTTCACCGCCCTTCGTTATCCAGTTCAGCGGGAACGCTC	2038		
Db	329	GACCGCTTCCACCGCGAACCGGTTCGCGCCCGGCGTACCCAGTTGCGCGGAAGGCGCTC	270		
QY	2039	CTG CACCGTTCGCGGTGGACAAAGCGCATCGGAACCGGCAATCGCGTTCTTCGATCGA	2098		
Db	269	CTCGCGCTTCGTCGCGTGCACCATCTCGATCCCCACGGGTGACACCTCGGCTCCGC	210		
QY	2099	GTTGCGCATCGGACACAGAAATGATCAAAACTACTGTTTCGGGCGATGGGTTCGCCCGAG	2158		
Db	209	GTCCATAGTCGGCACCGGATCGATCAAGACTGCTCTCGGGCATAACT-CCCCCGAGG	151		
QY	2159	ATCATGTGATCCGAGCTGTTTC-----TGTATGCGGCAACCTTAAAGATACAGACTGA	2210		
Db	150	GTCATGTAGTGGCAGCGGTTTCCTGCGATGCGCTACTGGTAGGACAAAGATACAGATAA	91		
QY	2211	CGCGTTCCTTTT 2221			
Db	90	TCGCTTTT 80			
RESULT 6					
SVU24659/c					
LOCUS	SVU24659	6252 bp	DNA	linear BCT 07-AUG-2003	
DEFINITION	Streptomyces venezuelae jad gene cluster, complete sequence.				
ACCESSION	U24659				
VERSION	U24659.2	GI:16445340			
KEYWORDS	Streptomyces venezuelae				
SOURCE	Streptomyces venezuelae				
ORGANISM	Bacteria; Actinobacteridae; Actinobacteriaceae; Streptomyces.				
REFERENCE	1 (bases 2899 to 6252)				
AUTHORS	Yang, K., Han, L., Vining, L.C. and He, J.Y.				
TITLE	Regulation of jadomycin B production in Streptomyces venezuelae				
	ISP5230: involvement of a repressor gene, jadR2				
JOURNAL	J. Bacteriol. 177 (21), 6111-6117 (1995)				
PUBMED	7592375				
REFERENCE	2 (bases 2899 to 6252)				
AUTHORS	Yang, K., Han, L., Vining, L.C. and He, J.Y.				
TITLE	Participation of jadR1 in the regulation of jadomycin B production				
	in Streptomyces venezuelae ISP 5230				
JOURNAL	Unpublished				
REFERENCE	3 (bases 2899 to 6252)				
AUTHORS	Yang, K.				
TITLE	Direct Submission				
JOURNAL	Submitted (11-APR-1995) Kegian yang, Biology, Dalhousie University,				



Db 2435 GAGCTGGTGGATGGGCGGAGCGCTCGCGCGCGGTGTTGACGAGGATGTCGAATC 2376  
QY 3522 GGGCGGCACTCGAACTCGCCATCCCGCGGTGGAACGCCGCGTAGAGCGCGCGCGTC 3581  
Db 2375 CCGCGCGCGCGCGCTCCGAGAGCGCGCTCGAGGCGCGGTAGAAAGTGTGACGCGC 2316  
QY 3582 ACCACGACGCGAGTTCGGCGCGGATGCCAACCCCTGTCGCGCTGCTCCGATGGT 3641  
Db 2315 GTCGGCGCGCGAGTTCGCGCGGAACGGAAGCGCGCGCGCCGCCCTCGATGAT 2256  
QY 3642 CTCGACGGTCTCTCGCGCGCGCGCTCGCTGTCGCGTAGTGAATGCCAGAGCGCCCC 3701  
Db 2255 CTCGACCGTCTCCAGGCGCGCTTCTCGTGTGCGGTAGTGAAGCGGACGAGCGCGC 2196  
QY 3702 GTCGCGCGCGAGCGGCGGATACCGGTGCGGATCCCGCGCTTCCCGGTGTCACGAG 3761  
Db 2195 GTCGCGCGCAAGTCTCGCGGATCGCTCTTCGATGCGCGGTTCGCCCTGTGACGAT 2136  
QY 3762 GCGGTCTTGCCTCCAGG-----GTCCTTCATACCTCGTCCCATGTGACGCAATAT 3814  
Db 2135 GCGGTCTTGC CGGCGGACGCGGACATGTGCGGTGTCCTACCTTCTTTCACGGAATGG 2076  
QY 3815 CAGCCCCCGGTGCGTGAGCGACCAATGGCGCGCTCGCGCGTTCGAATCGACGGTCA 3874  
Db 2075 TCGAGGAGTCCGTGTGTAACA CGCGCGTGTCCGCGAACCCCGTGAAGCCCGGCCCG 2016  
QY 3875 CAGCCTACTGTGACCGGTGACAGGGCGGAGTGCGCGGTGTTGAGCGGTGGGGCCA 3934  
Db 2015 GGTTCGACGCGAGTCCCGCGAGGCGGTGTCGCGGAAACAGTGTGCGTGGCGAGCA 1956  
QY 3935 GATCGGCGCGCGCACGGGNAACGGCGCGGTGAGGGGTGAGGGTCCGCGGACCGC 3994  
Db 1955 TGTCGAGTGTGCCCCCGAAGGGGAGGCGCGGAGCAGCCTGCGGGCGGTTCGGAGGT 1896  
QY 3995 CCA--GGCGGTGAGGCGACCGACCGATCGAGGTGCGCGCTGCACCGGCGCAACA-GT 4051  
Db 1895 CCACGCGCGTCCCCGGGGGAGGCCGTGCGGTTCAGCAGGACTCGCGGCGCGGAGGA 1836  
QY 4052 CCTCGCGCGCGAGTTCGACTC-GTACGCGTACGAGCGCGGTCCGCGCGAGTCTGAGC 4110  
Db 1835 GGTCCGCGACCGGCAACCGGTTCCGGATGTCGACGTAGTACGTTCCCGGCTCCGTCGCGG 1776  
QY 4111 TGGAGTGGCGCGCGGTGAGGCGGTTCGCGCGGCGGCGGAGGCGGAGGTGCG 4170  
Db 1775 GGGAGGCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1716  
QY 4171 GGGCGGAGCATCATGGGCGCGCGGTGCGAAACCGCGCGCGCGCGCGCGG 4217  
Db 1715 GCGAGGACCGCGGTGAGCGGTCCCGGCGCGCGCGCGCGCGCGCGCGCGCGG 1669

RESULT 7  
BA000030\_43/c  
WPCOMMENT

Sequence split into 91 fragments LOCUS BA000030 Accession BA000030

Fragment Name	Begin	End
BA000030_00	1	110000
BA000030_01	100001	210000
BA000030_02	200001	310000
BA000030_03	300001	410000
BA000030_04	400001	510000
BA000030_05	500001	610000
BA000030_06	600001	710000
BA000030_07	700001	810000
BA000030_08	800001	910000
BA000030_09	900001	1010000
BA000030_10	1000001	1110000
BA000030_11	1100001	1210000
BA000030_12	1200001	1310000
BA000030_13	1300001	1410000
BA000030_14	1400001	1510000
BA000030_15	1500001	1610000
BA000030_16	1600001	1710000

1700001 1810000  
1800001 1910000  
1900001 2010000  
2000001 2110000  
2100001 2210000  
2200001 2310000  
2300001 2410000  
2400001 2510000  
2500001 2610000  
2600001 2710000  
2700001 2810000  
2800001 2910000  
2900001 3010000  
3000001 3110000  
3100001 3210000  
3200001 3310000  
3300001 3410000  
3400001 3510000  
3500001 3610000  
3600001 3710000  
3700001 3810000  
3800001 3910000  
3900001 4010000  
4000001 4110000  
4100001 4210000  
4200001 4310000  
4300001 4410000  
4400001 4510000  
4500001 4610000  
4600001 4710000  
4700001 4810000  
4800001 4910000  
4900001 5010000  
5000001 5110000  
5100001 5210000  
5200001 5310000  
5300001 5410000  
5400001 5510000  
5500001 5610000  
5600001 5710000  
5700001 5810000  
5800001 5910000  
5900001 6010000  
6000001 6110000  
6100001 6210000  
6200001 6310000  
6300001 6410000  
6400001 6510000  
6500001 6610000  
6600001 6710000  
6700001 6810000  
6800001 6910000  
6900001 7010000  
7000001 7110000  
7100001 7210000  
7200001 7310000  
7300001 7410000  
7400001 7510000  
7500001 7610000  
7600001 7710000  
7700001 7810000  
7800001 7910000  
7900001 8010000  
8000001 8110000  
8100001 8210000  
8200001 8310000  
8300001 8410000  
8400001 8510000  
8500001 8610000  
8600001 8710000  
8700001 8810000  
8800001 8910000  
8900001 9010000

BA000030\_17  
BA000030\_18  
BA000030\_19  
BA000030\_20  
BA000030\_21  
BA000030\_22  
BA000030\_23  
BA000030\_24  
BA000030\_25  
BA000030\_26  
BA000030\_27  
BA000030\_28  
BA000030\_29  
BA000030\_30  
BA000030\_31  
BA000030\_32  
BA000030\_33  
BA000030\_34  
BA000030\_35  
BA000030\_36  
BA000030\_37  
BA000030\_38  
BA000030\_39  
BA000030\_40  
BA000030\_41  
BA000030\_42  
BA000030\_43  
BA000030\_44  
BA000030\_45  
BA000030\_46  
BA000030\_47  
BA000030\_48  
BA000030\_49  
BA000030\_50  
BA000030\_51  
BA000030\_52  
BA000030\_53  
BA000030\_54  
BA000030\_55  
BA000030\_56  
BA000030\_57  
BA000030\_58  
BA000030\_59  
BA000030\_60  
BA000030\_61  
BA000030\_62  
BA000030\_63  
BA000030\_64  
BA000030\_65  
BA000030\_66  
BA000030\_67  
BA000030\_68  
BA000030\_69  
BA000030\_70  
BA000030\_71  
BA000030\_72  
BA000030\_73  
BA000030\_74  
BA000030\_75  
BA000030\_76  
BA000030\_77  
BA000030\_78  
BA000030\_79  
BA000030\_80  
BA000030\_81  
BA000030\_82  
BA000030\_83  
BA000030\_84  
BA000030\_85  
BA000030\_86  
BA000030\_87  
BA000030\_88  
BA000030\_89

BA000030.90 9000001 9025608									
Continuation (44 of 91) of BA000030 from base 4300001 (BA000030 Streptomyces avermitilis)									
Query Match		8.7%	Score 377.2;	DB 1;	Length 110000;				
Best Local Similarity		64.1%;	Pred. No. 1.8e-24;						
Matches 568;		Conservative 0;	Mismatches 318;	Indels 0;	Gaps 0;				
Qy	227	GATCACCCTCGGCTTCGGCCGAGAGCTTCGCGTGTGTCGCGCGCCCGCCAGATGCGGAACGGC	286						
Db	62443	GAGCATGCCCGCATTCAGGCGAACGGCTTCCTGTGTGTCGTCGCCCACTGTCTCGTAGGAG	62384						
Qy	287	CTGGCGGGCGCGCCCGCTCACTTCCTCCGACGCGTCCGGCACGACCTGCGCTTGGGCCCCCGGCC	346						
Db	62383	CGTGGTGGGTGGCCCGTGACCCGCGGGAACGTGGGGTGGGTCTGTCGTCTTTCGCGCCGACG	62324						
Qy	347	CGCTGGCGCTCGGGCGCTCTCAGGAACGGCTCGCGACGGGCGCTCGGATACTTCCGGAGC	406						
Db	62323	AGCTGGCGTGTGTCGCTCAGGACGAGTGGTGCACCACTTGGGGCGGGTAGCGCCGCGCC	62264						
Qy	407	ATCTGCTCGCGCGCCCGCTCCAGCCCGCAGCTCTCTGAAACGCGAGTGACCGCCCGCAGCAC	466						
Db	62263	CACCGGGCGGACCTGCTGTGTGTGTCAGTTCGGCGCAGACGAGCGGCGCGCGAGGACC	62204						
Qy	467	TCGGAGAGCGCGCGCTGTGCTGCTGGCGGTGATGCGCTTCGGCGCCCGGACAGCGGTAC	526						
Db	62203	GCCGCCAGGTGCGCGCTCTGTGCTCCGCGGAGACCGCTTCGGGCGCGGTGAGGGGATAC	62144						
Qy	527	GCCGCTCTCTGTCGCGCGCGCGCGGTGAGTGCCTTGACCGCCACTTCCGCGATGTCGCGC	586						
Db	62143	ACCCGGCCCTCTGTGGCGCGGTGAGGTGAGCGCGTTCACCGCGACCTCGGCGATGTACGG	62084						
Qy	587	GGATCGACGCGAGCAACCGGGGACGTGCCGTACAGCGCGCGGACCAACGCGCTCGGACCGG	646						
Db	62083	GGGTGACGAGCGGCTTCTCCGAGCGCGGTACAGGGCTCGCACGTGCTTCCCGCGCG	62024						
Qy	647	ATGCGGGCGCCCAAGACAGCGTGTGGAATGATGAGTCTGCGCCCGCAGGAGGTCCAG	706						
Db	62023	ATGGAGTCCGCGCCAGGACAGGGTGTGGAATGAAGGCGCGGTTCGACGAGGCTCCAG	61964						
Qy	707	TCTAGCCCGGACTCGCGTAGCGGCGCGTCTGTTCTCGCGTCCGCGCGCGGTGATGAAGTCG	766						
Db	61963	TCGAGCCCGGAGAGCGCAGCGTCTCTCGCACTTCCGCTGCCACCGCGTGATCAGGTGCG	61904						
Qy	767	TCGCGCGCGCGTTCGCCCAACCGCGAGATGGAAGCTTTCACAGGTGCGCGACGCGCGCC	826						
Db	61903	TCGGCGTGTCTCTACCGCGCGGTACGAGGACAGCTTCAACACATGCCGTACGTTCGCGC	61844						
Qy	827	TCGCGCGCGCGCGCGGAAACGCTGCTGCTCGCGCTCGGTGGCACTGTTCTGTGACAGG	886						
Db	61843	GCCACCGCGCGCGCGGAAACCGCGCTCTCTCTGACCGCGCGGATCGACGCGTACCGAGG	61784						
Qy	887	AACGCGCGCGCACCCCGTTGAGGGCCGCTTCAGGCGCCCGGGCGGTTCGCGGTACTCGGCC	946						
Db	61783	AACGCGCGCTCGAGCGCCCTTTCAGTGCCTCGCAGCAGCAGCGGTGCTCTGTACGCTCCG	61724						
Qy	947	GCGCAGACCTCGAGCTTCGGGCGGTGACCGTTCACCGTTCGCGCCCGCGCGGAGGACT	1006						
Db	61723	GTGGCGACCTCGACGCGCGGCGCTTTCACCGTTCGAGCGCGCGGTTCGCGGACCGAAGG	61664						
Qy	1007	CTGACGGGACCGGTCGCGGCGCAGCAGGTGGCGGACCTTGAAGCGCGCGCACACCGGTACG	1066						
Db	61663	CGCACCGCGCGGACCGCGCCAGTGTGTGCGCACTTCGCGCGCGCCCAACACGCGGTTCGCG	61604						
Qy	1067	CCGGTCAAGAAATCACTCGGGGCTCTCTCTCGGCGAGCGGACGCG 1112							
Db	61603	CCGGTCAAGAAATCAACCGCACAGGCGCTTTCGTCTCGGGCGACGCG 61558							

RESULT 8  
AY228176S2/c 5495 bp DNA linear BCT 02-APR-2003  
LOCUS Streptomyces WP 4669 PD 116740 angucycline type II polyketide  
DEFINITION synthase gene cluster, partial sequence.

ACCESSION	AY228177	GI:29469247
VERSION	AY228177.1	
KEYWORDS	2 of 2	
SEGMENT	Streptomyces sp. WP 4669	
SOURCE	Streptomyces sp. WP 4669	
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.	
REFERENCE	1. (bases 1 to 5495)	
AUTHORS	Yin, X.H., Yoon, E., Mahadevan, B. and Proteau, P.J.	
TITLE	Molecular cloning and sequence of the PD 116740 angucycline type II polyketide synthase gene cluster from Streptomyces WP 4669	
JOURNAL	Unpublished	
REFERENCE	2. (bases 1 to 5495)	
AUTHORS	Yin, X.H. and Proteau, P.J.	
TITLE	Direct Submission	
JOURNAL	Submitted (31-JAN-2003) Department of Pharmaceutical Sciences, College of Pharmacy, Oregon State University, Pharmacy Building Room 203, Corvallis, OR 97331-3507, USA	
FEATURES	Location/Qualifiers	
source	1. .5495	
	/organism="Streptomyces sp. WP 4669"	
	/mol_type="genomic DNA"	
	/strain="WP 4669"	
	/db_xref="taxon:224730"	
	/note="adjacent to AY228176; approximately 50 bp gap between adjacent sequences due to extremely high GC content"	
gene	50. .835	
CDS	50. .835	
	/gene="pd2E"	
	/gene="pd2E"	
	/note="pd2E"	
	/codon_start=1	
	/transl_table=11	
	/product="putative ketoreductase"	
	/protein_id="AA065365.1"	
	/db_xref="GI:29469255"	
	/translation="MSDKHERVALVSGATSGIGLAAARLLAQGHVFLGSRDADNVA RTVKELTABGLDVGDTLDVRSALSIESFVQAAVDRFGVDVLVNAGRSGGVTADI DELWYDVETNLNSVFLRTAVLTGGMKRNRIINIASTAGQGVVLGAPYSAS KHGVGPTKALGNELAPGTITVNAVCPGVETPMQVRVAGYAAAYDTTDAIMEKFT AKIPLGRYSTPEEVAGLVGLASDTAAASITAAQLNVCGLGNF"	
	877. .1821	
	/gene="pd2D"	
	877. .1821	
	/gene="pd2D"	
	/note="pd2D"	
	/codon_start=1	
	/transl_table=11	
	/product="cyclase-like protein"	
	/protein_id="AA065366.1"	
	/db_xref="GI:29469256"	
	/translation="MSASELREVEHEITIGAPADVRLADVANNPRIFPTIYVDH LERSDESLIRIWTANGEPKNWTSRRVLDREALRITFQQVSTPPVATMGSTWILEP QGDGASRVLLHDYTDVDDPAGLKWIDVARNRSSELASLTKNLVLAASABLTFS FEDTVYDGSQADVYDFVNEAGLWSERLPHVATVRLTSDSPGLQLEMDTRADGSGTH TTSRVCLQCKIAYKQITLIPALMLHTGYWTFDTDTGTSTASSQHTVVLRENIEK ILPGGAVAREYVRSALSTNSRATLGHAKAYAEARR"	
	1824. .4118	
	/gene="pd2OR"	
	1824. .4118	
	/gene="pd2OR"	
	/note="pd2OR"	
	/codon_start=1	
	/transl_table=11	
	/product="oxygenase reductase-like protein"	
	/protein_id="AA065367.1"	
	/db_xref="GI:29469257"	
	/translation="MAGADAATVDVVVVGAGVGLMAGELRGLGADVTVLRDLPTPT TESRASTLHARTMELLKORGLDGDGDFDPLPHEPRGHFGGLPMDLTLPGSPYQGM KYQVTRTELLGARAARLGVRLRGHEVTGLTQDQGVTSARHAGRPVRFRTARYVG CDGNSAVRRRLAGVDVDPGDDAERELIRADVGGIDIPNRRFQRLPVGLAIARRGDGT	

RVVHRTCTPRDEARFADIVEAWREVTGEDIAGHTPLWVNSFGDASROAARVVRGR  
VLLAGDAHQOMP1GGQALNLGLQDAVNLGWKLAAATVRGAPEGLLDTYHDERHVGPR  
RVLSNIRQAQRLLLGGPEVEALRSYIGELTSYERVRAHLAGTISGLDVRHGTGCGPTL  
AGTLRPAALAGTTPRSTDLRLREARGLLLVLTDAPEPFARVARPWSGAVDTVAARP  
AEAGLPEAAALVRPDGYLAWAGDTPEGLDAALRWFTPGPTPAPRHTPHRHQHVYR  
RTKMGNTLRTALVTGASRGMGTAORLARDGALVAHVYGTNEDDAAEVAVAGIERDG  
GRAFAVRAELGAGDVHLEFLALRGKERTGSTELNLLVNNAGVMGCVKPEDTTPDM  
FDRI FAVNARAPPFLVQBALTNMFDGGRI VNISSGLTRPAMPDEIAVMSKGAVEOLA  
LHYKALGPRGITVNSVAPGITRNGNALFDIPEAVEAVALUSTFNRVGEPODVADVVA  
FLVSDARWVTGSGFVDSGGTLLG"  
4123...>5495  
/gene="pd2j"  
CDS 4123...>5495  
/gene="pd2j"  
/note="pd2j"  
/codon\_start=1  
/transl\_table=11  
/product="putative transporter"  
/protein\_id="AAO65368.1"  
/db\_xref="GI:29489258"  
/translation="MTRTHPAPVAGADRWSPRLWGLLLVLVAGNMLIDALEVSVAVV  
ALPSIGTDLGPLELTDVHWMTGFVAGFGGLLFGARVVSLLGRVRRVLAALLAFAAS  
ALSALAGPEALLVATRLVKGFCAALTAFTGLAIATAPQGPARDRAVSVVAFGAGS  
FTAGLLSGMLVTEYSWRLAFAPFAPVVLVLPALALRLIPDEDPAPARTSPRYDLTGA  
ATFETAALIALVYIGAVPAHGWTSFRAAGPLTAAVLLAVFAVVERTTSAFLVOPAAAR  
RPPLIRSLAGACLNGSVLGLLFLVITYOMOTVAGWPBSTAALALLPAGPPALATALP  
GRVAFGAPGLIAGAACAFAFGLQLSPAWPLHYPARLLPTLLLVGAAPVLSFAAL  
NQATSGLSPPERRGAAGQYQTVOLAAVVVAVALAAAAEPGATAAFAARRGLUPA  
GADPGGRA"

ORIGIN

Query Match 8.0%; Score 347.2; DB 1; Length 5495;  
Best Local Similarity 56.7%; Pred. No. 2e-21;  
Matches 773; Conservative 0; Mismatches 563; Indels 28; Gaps 6;  
QY 2942 CGTGCCCGAGCGCGACCGGGCGCGCTACGGGCGCGCGCGCGCGCGGTAGTCTGCC 3001  
DB 4201 CGAGCAGAGCGCCCCACAGCGCGGGGCTCCAGCGCCCGCTCGCGCGCGCGAGCGGGCG 4142  
QY 3002 CTGCGTACCGAAGCGTGGCGG----GTGAGAGATCGTTCCGCTGTGGCATTCGACGTAC 3057  
DB 4141 CGGGGTGCGTCTGTCGTACGGCGTCAACCGAGGAGGGTGC CGCGCTCGACGGAAG 4082  
QY 3058 TGGCGGTGATCCACCGTGAAGTCTCGAGGAGCGAGAGAGGCGCACCGTGGCGGATGTCG 3117  
DB 4081 GAGCCCGTGACCCAGCGCGCGTCTGTCGAGAGCGAGGAGGCGACACGTCGCGGACGTCC 4022  
QY 3118 TCGGCTCTCGCGATCGCGGTGAAACACGAGTGTGGCGGCGAGTCCGCGCGCGCTCGGGG 3177  
DB 4021 TCGGTTTCGCGACCCGTTGAAGTGAACAGGGCGCGCATCGCTCGACGCGCTCGGGG 3962  
QY 3178 GTCTCGCGCGCGGTGCGTTCATGTCGTCTTCACGAAACCGGCGCGCACCGCGTTGACC 3237  
DB 3961 ATGTGCAACAGCGGTGCGGTGCGGT-----GATGCCGGGGCGACGCTGTTGACG 3908  
QY 3238 GTGATCCCGCGTTCGCCAGTTGCTTGGCGAGGGCGAGGTGAGCGTGTCCACCGCACCC 3297  
DB 3907 GTGATCGCGCGCGGCGCGAGGCGCTTGGCGTAGTCAGCGCGAGCTGTTGACGCGCGCC 3848  
QY 3298 TTGGTTCATCGGTATCCGATGAGCTCGGGGACGCGCGCGGTGCGCGGACGAGATG 3357  
DB 3847 TTGCTCATGGCGTAGGCGATCTCGTCTGGGTGTTGGCGAACCGGGTCAGGCCCGAGAGATG 3788  
QY 3358 TTGATGATCCCGCGCGCGTTCGCGAGTCTGTTTTCAGTCCGTGTGGACGAGGAACGCGGT 3417  
DB 3787 TTCAGATCCCGCGCGCGTTCGCGCATGTTGCTGAGCGGCGCGTGGACGAGGAACGCG 3728  
QY 3418 GCCCGGAGCTTGAACGCGGACGAGTCTGTCGAGAGACTCTCTCGGTGACTTCGTTGATCCGT 3477  
DB 3727 GCCCGGGGCTTGACGCGCGAAGATCCCGGTTCGAACTATGTCGCGGTGTTGCTTCGGGCTTG 3668  
QY 3478 CCGGAGCGCGTACCGCGCGTGTGTTTACAGGATGTTCGAACTTCGCGGGCGCACTCCGAAC 3537

DB 3667 ACCCGCCCATACAGCCCGGGGTTTGTTCACAGGATGTTACGTCGTCGAGCCGCGTCCG 3608  
QY 3538 TCGCCCATCCGGCGTGAACAGCGCGCTAGAGCGCGCGCGCTCACCCACGACGCCGAGT 3597  
DB 3607 TCCTTCAGGCGCTGCTCCAGGGCGAGAAAGCTCTGTCGACAGTCCCGGGGGCCCGAGT 3548  
QY 3598 TCGCGCCCGATGGCAACAGCGCTGTCGCCCGCTGTCTCCGGATGCTCTCGAGCGTCTCTCGC 3657  
DB 3547 TCGCGCGGTACGGGCAACAGCGCGCGCGCTCAGCTCGATCCCGGCGACGACCTCGCG 3488  
QY 3658 GCGCCCGCTCGTGTCTGCTAGTGAATGTCACAGAGCGCGCTCGCGGGCGAGCGCG 3717  
DB 3487 GCGCGTCTCTGTTGTCGCTAGTCAGCGCGACACGAGCGCGCTCGCGGGCGAGCGCG 3428  
QY 3718 AGGGCGATACGGCTCGATGCCCGCGCTTCCCGCGTTCACAGGGCGGTCTTCGCCCTCC 3777  
DB 3427 TCGCGGTGCGCGCGCGCATGCTCGCTGGCGCGCTGAGAGCGCGCTCTCCCGGTC 3368  
QY 3778 AG-----CGGTCTTCCATACCTCGTCCCATGTGACAGCATATACAGCCCGCGCG 3826  
DB 3367 AGTTTGCCCATCTTCGTCTCTGACGTGGTGTGTCGTTGGGATGCGTGGGTGCGGT 3308  
QY 3827 TGCGTGAGC-GACCATGGGGCGCGCTCGCGCTTCCGATTCGACGCTCAGACCTACCTG 3885  
DB 3307 CGCGTGGCGCGACCGCGCGTGAACACAGCGCGCGAGCGCGCTCGAGTCTCTTCGGGGTG 3248  
QY 3886 TGACCGGTTCAGAGCGGGCGCGAGTGGCGCTTGGAGCGGTGGGGCGCAGATCGGGCGCG 3945  
DB 3247 TCGCGCGCGCCAGGCGAGGTATCCGTCTGGGGCGTACGAGCAGGGCGCGCGCTTCGGGCG 3188  
QY 3946 GCGCACG----GGGAACCGCGCGCGTTCAGGGGTTCAGGGGTTCGCGGGGACCGCCAGGCC 4001  
DB 3187 GCGCGCGCTCGCGCGGGCGCGCTACGCTGTCGAGCGCACCGGACCGCGCGGGCG 3128  
QY 4002 GGTTCAGGGCACCG--ACCGGATCGAGTTCGGGCGTGCACGCGGCGCACAGTCTCTCGCG 4059  
DB 3127 ACCCTGGCGAACCGTCTCGCGCGCTCGTTCAGACACGAGCAGCAGCGCGCGCGCTCGCG 3068  
QY 4060 CCGAGTCTCGACTCTGACGGGTACAGAGCGCGGTTCGCGCGGAGTCTTGAGTTCGAGCTGG 4119  
DB 3067 AGCAGGTTCAGGGTGGAGCGGGGTCCGCTGGCGCCAGAGGGCGGGCTTCGGCAACCG 3008  
QY 4120 CCGCGCGGTGGGTGAGCGGTTCGCGCAGGGGCGGAGAGCGGGGAGGTTCGCGCGCGCAGC 4179  
DB 3007 GTCCGCGGAGCTGGGTTCGCTCGCGGTGCGTGGCGAGCTCCAGGCGCTGATCGT 2948  
QY 4180 ATCATGGGGCGGGCGCGGTGAAACGCGCGCGCGCGGTTCAGAGGGTCTCTCCAGGAGC 4239  
DB 2947 CCGCGCAGGTGCGCGCGCACCGCTCGTACGACCTTCAGTTCGCGGATCACCGAGCGCAGC 2888  
QY 4240 TCTAGACCGCGCAACCGCGCTTCGCGCAGCGGGCGAGCGCGCG 4283  
DB 2887 GCCTCGACCTTCGCGCGCGCCAGCAGCGCGCGCTTGGGGCCG 2844

RESULT 9  
SC0939127 290850 bp DNA linear BCT 16-APR-2005  
LOCUS Streptomyces coelicolor A3(2) complete genome; segment 24/29.  
DEFINITION AL939127 AL023496 AL023861 AL031155 AL031182 AL031232 AL035161  
ACCESSION AL035205 AL035206 AL132824 AL512902 AL590982 AL591083 AL591084  
AL645982  
AL939127.1 GI:24429552  
VERSION Streptomyces coelicolor A3(2)  
KEYWORDS Streptomyces coelicolor A3(2)  
SOURCE Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
ORGANISM Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1  
Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L.,  
Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,  
Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,  
Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S.,



Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S., Rabinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S., Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S., Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G., Parkhill,J. and Hopwood,D.A.  
Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)  
Nature 417 (6885), 141-147 (2002)  
12000953  
2 (bases 1 to 290850)  
Bentley,S.D.  
Direct Submision  
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sds@sanger.ac.uk  
On or before Oct 30, 2002 this sequence version replaced  
gi:20520661, gi:20520749, gi:20520754, gi:20520755, gi:20520688, gi:20520689, gi:20520815, gi:20520683, gi:20520917, gi:20520686, gi:20520864.  
Location/Qualifiers  
1..290850  
/organism="Streptomyces coelicolor A3(2)"  
/mol\_type="genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
97..1101  
/gene="SCO6257"  
/note="synonym: SCAH10.22"  
97..1101  
/gene="SCO6257"  
/note="SCO6257"  
/note="SCO6257"  
binding lipoprotein, len: 334 aa; similar to TR:CA841563 (EMBL:AL049727) Streptomyces coelicolor putative secreted solute binding protein, 337 aa; fasta scores: opt: 1003 z-score: 1130.9 E(0); 49.8% identity in 325 aa overlap, z-score: 1130.9 E(0); 49.8% identity in 325 aa overlap, to SW:RBSB\_BACSU (EMBL:Z32953) Bacillus subtilis D-ribose-binding protein precursor RbsB, 305 aa; fasta scores: opt: 206 z-score: 238.7 E(0); 6.3e-06; 23.0% identity in 235 aa overlap and to TR:CA841563 (EMBL:SC9B1) Streptomyces coelicolor SC9B1.17c, 337 aa; fasta scores: opt: 1003 z-score: 1045.2 E(0); 0; 49.8% identity in 325 aa overlap. Contains match to Pfam entry PF00532 Peripla\_BP\_like, Periplasmic binding proteins and LacI family and a match situated in the correct position to Prosite entry PS00013 Prokaryotic membrane lipoprotein lipid attachment site. Contains also possible N-terminal region signal peptide sequence"  
/codon\_start=1  
/transl\_table=1  
/product="putative ABC transport system sugar binding lipoprotein"  
/protein\_id="CA860176.1"  
/db\_xref="GI:6273662"  
/db\_xref="InterPro:IPR001761"  
/db\_xref="UniProt/TREMBL:Q9RKT5"  
/translation="MARFTWGIAGALGSLVSLAGCSSTGKRAEDARKAASRGRA AVDTPRWTAMITHSGDGTFDWIVQSGAEQAAVDNINFLYSHDAEQOQLVNAALDKKVDGIVTLAKPDMSALABAHKAGIPVITVNSGSESKERFALTHVQDQETIA IGKVAEELNREQRQAVCLVHQGNVGHQSCDGVKERTFDGKQRLVYNGTSPDVQS ATEAKLQTDSDVTVTGAPYADTAVKAKQAGSKAEIDITFDLNKAAVAGLADGLTG FAVDQOPLYQGYEAVDLLWLYKNADVLGGSPVLTPGQIITKDDAALADYTKRGTR"  
133..165  
/misc\_feature  
/gene="SCO6257"  
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"  
241..966  
/misc\_feature  
/gene="SCO6257"  
/note="Pfam match to entry PF00532 Peripla\_BP\_like, Periplasmic binding proteins and LacI family., score 13.80, E-value 2.1e-06"  
1098..2138  
/gene  
/gene="SCO6258"

/note="synonym: SCAH10.23"  
1098..2138  
/gene="SCO6258"  
/note="SCAH10.23, probable ABC transport system sugar permease, len: 346 aa; similar to TR:O68120 (EMBL:AF010496) Rhodobacter capsulatus ribose transport system permease protein RbsC; fasta scores: opt: 445 z-score: 508.0 E(0); 6.4e-21; 30.5% identity in 351 aa overlap and to SW:RBSB\_ECOLI (EMBL:L10328) Escherichia coli ribose transport system permease protein RbsC, len: 321 aa; fasta scores: opt: 286 z-score: 329.9 E(0); 5.3e-11; 30.1% identity in 329 aa overlap. Contains possible hydrophobic membrane spanning regions"  
/codon\_start=1  
/transl\_table=11  
/product="putative ABC transport system sugar permease"  
/protein\_id="CA860177.1"  
/db\_xref="GI:6273663"  
/db\_xref="GOA:Q9RKT4"  
/db\_xref="InterPro:IPR001851"  
/db\_xref="UniProt/TREMBL:Q9RKT4"  
/translation="MNATQVRAADERILQTSRLKRLGRPELGSVVGAIIVFPFAFF ADSFLHAASTLVLYAASTIGAFNAPVALLMIGGFDSAGVMVTSALVSMFSYQM TANYVGVVSLVTLAIGAFNGFMLTRKLPSFIITLTGLMLTKGVKVDGT VSTKSIADMEGFPSSAODVFASITIGVGFKVITLMLVALVAVSWIGLRTAGNWIF AVGNKQDARAVGYPVAKTKIGLYMGVGFAGWISQHLLEFSDVVDVQSGEGVGNELIYI IAAVIGGLITGVGSAVGSVAVGAFIFGTSKGIYFAEWNPDPWFKEFLGMLLLATLL NAWVRKRAEATK"  
2124..2127  
/gene="SCO6258"  
2135..2926  
/gene="SCO6259"  
/note="synonym: SCAH10.24"  
2135..2926  
/gene="SCO6259"  
/note="SCAH10.24, probable ABC sugar transport ATP binding protein, len: 263 aa; similar to TR:O68121 (EMBL:AF010496) Rhodobacter capsulatus ribose transport ATP-binding protein, 305 aa; fasta scores: opt: 612 z-score: 706.1 E(0); 5.9e-32; 45.6% identity in 217 aa overlap, to SW:RBSA\_ECOLI (EMBL:M13169) Escherichia coli, ribose transport ATP-binding protein RbsA, 501 aa; fasta scores: opt: 509 z-score: 585.6 E(0); 3e-25; 31.5% identity in 254 aa overlap and to TR:CAAL15787 (EMBL:SC7B7) Streptomyces coelicolor St7B7.07, 260 aa; fasta scores: opt: 592 z-score: 521.2 E(0); 1e-23; 43.1% identity in 246 aa overlap. Contains pfam match to entry PF00005 ABC tran, ABC transporter and Prosite PS00017 ATP/GTP-binding site motif A (P-loop) and PS00211 ABC transporters family signature"  
/codon\_start=1  
/transl\_table=11  
/product="probable ABC sugar transport ATP binding protein"  
/protein\_id="CA860178.1"  
/db\_xref="GI:6273664"  
/db\_xref="GOA:Q9RKT3"  
/db\_xref="InterPro:IPR003439"  
/db\_xref="InterPro:IPR003593"  
/db\_xref="UniProt/TREMBL:Q9RKT3"  
/translation="MTRNEDTALVELSVSKYGNVRALGVSLVHAGEITCVLGD NGAGSTLTKIAGLHQHDDGGLSDGBETRLSSPREALDRGIATVYQDLAVVPLMPV WRNFFLGSPPRGKGVAPFKRMVDVHRRTHAELRLMDLRDVPDIPQIVLAVVPLMPV AIFARAVYFGAKVLVLDPEALGVQSGVLKYAAARDQGLGVLLITNPHHAYLVG DEFVLLKSGTWVANQTRDEVTLDELTTQMGAGTGELDDLRLHELRG"  
2240..2815  
/gene="SCO6259"  
/note="Pfam match to entry PF00005 ABC tran, ABC transporter, score 168.70, E-value 9.6e-47"  
2261..2284  
/gene="SCO6259"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
2588..2632  
misc\_feature  
misc\_feature  
misc\_feature





```

gene
/db_xref="taxon:1894"
complement(416..1033)
/gene="aur18"
CDS
complement(416..1033)
/gene="aur18"
/note="putative transcriptional repressor"
/codon_start=1
/transl_table=11
/product="Aur18"
/protein_id="AAX57186.1"
/db_xref="GI:61968687"
/translation="MIKQRAVTRRALIESAARVGRGAYEATLSMISVGAVSPG
ALHEHPENKAAVAAVEAAARALSAAREVYARRTSALQALDSHALAGLLSDTV
ARAGQLSEPTPSYALPGEMWYVHRLAABAHEDALLPLNRQSVATTVAATL
GFEVLGRDDPQWLAPRTLAFGRVVMPCLAAPATLFLGAAGRS"
1192..1197
1216..1221
1261..12031
/gene="aur18"
1261..12031
/gene="aur18"
/note="response regulator-like protein; transcriptional
activator essential for auricin biosynthesis"
/codon_start=1
/transl_table=11
/product="Aur18"
/protein_id="AAX57187.1"
/db_xref="GI:61968688"
/translation="MNQRTMAVLEPHTVAAGNGSPAGGPRTPQHVPQAQALWEIL
IVESNSAEATLARGRHGHEVDIVGTGGEALRYVAEDLVLLDLPLDGLGVCR
GIRAHEIPVIAVTAGSELDRLVGLQAGADDYLAQPGFRELMAWMAVRRARPR
PAFATITCGPLSDGSTRVTLTGVEVTRKEFDMLYLLASHDPTVIPKMLMOQVW
GDSWSRRVTDVTHVSTLRNKIGAWIITRGVGRFG"
2099..3001
/gene="aur10"
2099..3001
/gene="aur10"
/codon_start=1
/transl_table=11
/product="aur10"
/protein_id="AAX5995.2"
/db_xref="GI:62177082"
/translation="MTVTALCHGESCRGADATGTRRRRAADRGSRLCRCRCHLAAGL
GEVAHYEGCGRVILGAGPGSLKERTTGSESTGLPENGAAADVARMVSVLGSWGLV
AQOGLTAPPTVADLTAFLLRNADWLAHPAAGVCEVARLVRAGRRAAPPEPART
FRIGTCPEPCGALALSVRGAERDRAVCDADPAHRTSDRWSELSDSLESGRAP
GAVGVPVAPAPAPATERNLTAADIAQLRHTPTGTVYRLASEERWRRNRAGRTFY
AEGDVACFAAGRRPAPPPPHR"
3170..3175
3193..3198
3250..4716
/gene="aur1A"
3250..4716
/gene="aur1A"
/note="oxygenase-like protein"
/codon_start=1
/transl_table=11
/product="aur1A"
/protein_id="AAX57188.1"
/db_xref="GI:61968689"
/translation="MDAPVTVVAGPAGMMLAGELRAGVAVIVLERLERRTGESRGL
GFTARTLETFDQIGILRFGDIERSTLGHFGPLDPLGVDGCAQAAKTPVQSVTETH
LEEWATGLGADFERGHEVLSVDGDTVAVEVRPGGTHTLRADMLVCGDGGRTVRK
AAGDPFGTAATMEMFLADIKGIELKPMIGETLPGQVMVWGLPGSTTRIVVCERG
PPRRRAPPFEEVAAWQRLTGEDI SAEPVWVSFAFGDARQVTEYRGRVLLAGDA
ATHLPAGQGNMTGIQDANVLGWKLASVVLGRAPLALLDTHYGERHPVGRLLMNR
AQGLFLSGAEVQLRDSRELWYEDVARHAAWVTGLEITTYDVGGKGPLGRMP
GLELAGFGRTSAALLCPARGVLLDLADNPRLRQRAAAWGDVRVDTVTARPAGAAGG
LPDPTTALLRPPGYVAAWAPGSHDLPMALRWFGPA"
4891..5538
/gene="aur1B"
4891..5538
/gene="aur1B"

/note="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="Aur1B"
/protein_id="AAX57189.1"
/db_xref="GI:61968690"
/translation="NATPLATIERAPSGPAGRRREIVRAARTVLAEGWTRCTABATA
AEAGVTADTVALLFPDHEQLLGLVLLDSSASVAALAAVADSHLEEVTLDEADLIAG
RAWLAPLEAFREHFALVRHSABVTRLPTEAVDKWQATGTPQRRRLARLRRLASMG
LLDISDAQAEERFILLVPSGVVQSRFLGALPLRDEADALLADGVADFIRLYRP"
5621..5953
/gene="aur1C"
5621..5953
/gene="aur1C"
/note="cyclase-like protein"
/codon_start=1
/transl_table=11
/product="Aur1C"
/protein_id="AAX57190.1"
/db_xref="GI:61968691"
/translation="MTHSNLIVARMDSSSDQVAKLFTBEDKTDMPGRMGTLRRLRFM
YKGLYFHLQDFAEADDGQRIQESRTDPRFMRISHDLLPFIQYDPATWRTPADAMATR
FYDWGASE"
5991..7259
/gene="aur1D"
5991..7259
/gene="aur1D"
/note="B-ketoacyl-ACP synthase-like protein"
/codon_start=1
/transl_table=11
/product="Aur1D"
/protein_id="AAX57191.1"
/db_xref="GI:61968692"
/translation="MTRRVITGIGVLAPGGIGTTAFMNLSSGRTATRGITLFDPA
FRSVAEVDPFPERHLSLEVRMDRAAQFQVWATREALADSGLDLAFDPYRIGI
AMGTGVAISGLDTRVRSADSEGRVIAQVDMYAPGLHYDHFVSPSFAVEAVAGAE
PSTVSTGCTSGIDSVAHAVDLIRGSADIMTGTADAPISPLACFADAIKATTALN
EPEPTASRPFDAIRKGFVLGGAAPFVLEELGAARRGRGAHYAEIAGYASRSRPFMT
GLRPDRENAEAITVALDEARLDPGDLDVYNAHSGTRQNDLHETAAAYKLALGPHHR
TPVSSVKMIGHISLGSVSGMEIAASVLAHMEHVVPPTANLRTDPDCELDYVPLVARD
HRTDAVLTVSSGGFGQSAMWLARPDRTA"
7256..8470
/gene="aur1E"
7256..8470
/gene="aur1E"
/note="chain length factor-like protein"
/codon_start=1
/transl_table=11
/product="Aur1E"
/protein_id="AAX57192.1"
/db_xref="GI:61968693"
/translation="MSARILVTGIGVAAPSGLGVDFWSVTRIGKNAIGPVTRFDASA
YPSRLAGEITHGFEKHELPQLRVPTQDRVLTQALVAADCAFDAGIEPGTIDPYAMGV
VTAGAGGFPAENELKLMSEGAHVSAQSVAFVAVNSGOISIRNGLRGAGVVI
SDQAGGDLAQAARQLRKGLIATGGDAPICSLGNASHLGGMLMSDEPERAVL
PFDDAAAGVYPGEGGAMLLEDEDSARDGARTVYGFAGYGAITLDPKPGSGREPLR
RAIDVALTDAACHPAEDVVFADGRHPLRDLREAEAITRVFGPRAVPTVPTKMTGRI
NSGGAPIDVVSALVSNREGLIPTTNVELSDAYDLDLVAVRPRPTASVRTALVLRGR
GFNSAVVVRVD"
8568..8852
/gene="aur1F"
8568..8852
/gene="aur1F"
/note="acyl carrier-like protein"
/codon_start=1
/transl_table=11
/product="Aur1F"
/protein_id="AAX57193.1"
/db_xref="GI:61968694"
/translation="MKGSLISAPEFTFDLKRILLLEGAGDEGVLDGDLDSSEFETL
GYESLALLETGGRIEREFGITLDDVDFTDNPTPRSLVTGINESLSALVDA"
8946..9731
/gene="aur1G"

```

CDS	8946	9731	/gene="auriG" /note="ketoreductase-like protein" /codon_start=1 /transl_table=11 /product="AuriG" /protein_id="BAK57194.1" /db_xref="GI:61968695" /translation="MSDDQKRVAVVTGATSGIGLAVARLLAAQDHRVPIGARSAENVA QTVKELQDEGLEVDGAVVDRSTGVRAMIQAADVRFSGVDVNVNNAAGSGGVTADI
Query Match	7.4%; Score 322.8; DB 1; Length 17455;		
Best Local Similarity	53.9%; Pred. No. 1.6e-19;		
Matches 713; Conservative	0; Mismatches 597; Indels 12; Gaps 2;		
QY	2896	CGGGAAGGACTGACCGCCGACGCGCCACCGGATACCGACCCCGCTGCGCGAGCGCGC	2955
DB	13186	CAGGAGGATCGGGACCGGGGACCGGAAGCGAAGAGGGGTGGGCCCGACCGGAGGGTT	13127
QY	2956	CGACCGGGGCCCTACGGGCGCGCGGGCGGGCCCGTAGGTCTGCCCTCGCTA	3009
DB	13126	CCCGTACGGCCACCCCTCGCGCGCGCGCGGGGTAGCGCCGCGGAGTCCAGAGCGGG	13067
QY	3010	CGAAGCGTGGCGGTACAGAAATCGTTCGCGCTGTGGCATCGACGTACTGGCGCGGTGATC	3069
DB	13066	CGGACAGGGCTACCGCGAGAGCGTGCGCCCGTGGGTGCGACGAAGAACCGGTGATC	13007
QY	3070	CACCGTGAAGTGTGCGGAGCGCAGAAAGGCCACACGTGCGCGATGCTGCGGGTCTGCCG	3129
DB	13006	CACCGGGCTGCTGCGTGGCGAGGAACGCCACACGTGCGCGACGCTCTCGGTTTCGCCCC	12947
QY	3130	ATGCGGTTGAACACGAGGTGGCGGCGCAGTGCCGCGCGCGCTCGGGGTCTGCCCGCGC	3189
DB	12946	ACCGGTTGAAGGCGGACAGCGAGGCGCATCTGCTCGACCGCTTCGGGGATGT-----CG	12893
QY	3190	CGTGTGTTTCATCTCGTCTCCACGAAACCGCGCGCACCGGTTCACCGTGATCCCCGT	3249
DB	12892	AAGACGGGGTTCCCGTTGGGGTGATGCGCGGCGCGACGAGTTGATGTGATGCCCCCG	12833
QY	3250	TCCCCAGTTGCTGGCCAGGCGGAGCGGTGTCCACCGCACCCCTTGTCATCCG	3309
DB	12832	GGGCCAGCGCTTCGCTAGTGCAGGGGAGCTGTTCGAGCGCCCTTGCTCATCCG	12773
QY	3310	TATCCGATGGAATCGGGGAAACCGCGCGCGGGTGGCGGACGACGATGTGATGATCCGC	3369
DB	12772	TAGGCGATCTCTCGGGTTGGCGAAGCGGGTCAGGCGGAGGAGATGTGATGATCCGG	12713
QY	3370	CGCGCGTGGCAGTCGTTTTCAGTCCGTGCTGGACCGAGGACAGGGTCCCGGACGTTG	3429
DB	12712	CGCGCGTGGGCGATGTTCCGACGCGCCGCTGGATCGAGAAACCGGCGCTTCGCGTTG	12653
QY	3430	ACGGCGACAGTCGGTTCGAAGACCTCTCGGTGATCTTCGTTGATCCGTCGCGAGCGCTG	3489
DB	12652	ACCGGAGATCCGTCGAACTGGTGGGATGATGTCGGCGGCGGACGCCCCCATC	12593
QY	3490	AGCGCGGTTGTTTACCAAGATGTCGAACCTCGGCGGCGACCTCCGAATCGGCCATCCG	3549
DB	12592	ACCGCGGCTGTTTACCAAGATGTCGAAGTGTGCTGCGCGTGGCTCCAGCAGGTCC	12533
QY	3550	GGTTCGMAAGCGCGGTAGAGCGCGCGCTCACCGACGCGGAGTTTCGGCCCGGATG	3609
DB	12532	GTTTCAGGCGCGGAAACAGGCTCTCCAGTCTCGGGGACCGCGAGTTTCGGCCCGATG	12473
QY	3610	GCCAAAGCGCTGCGCGCTGCTCGGATGCTCGACGCTCTCTCGCGCGCGCCCTCG	3669
DB	12472	GCGAAGGCCCGCGCGCTCTCTTCGACCGAGGCGGACGACCTCTTCGCGCGCGCCCG	12413
QY	3670	CTGCTGCGCTAGTGAATGCCACGAGCGCGCGCTCGCGCGCGCGCGGACGAGCGGATACG	3729
DB	12412	TCACTGGTGTAGTGGACGCGCGAGCGCGCTCTCTCGCGCGCGCTGATCGCGCGCGG	12353
QY	3730	CGTTCGATGCCCGGCTTCCCCCGGTTCACGAGGCGGTCTTCCCTCCAGCGGTCTTCCA	3789

Db

12352

CGGCCCATGCCCGCTCGATCCGTTGACAGCGCGGTTCGGCGCGGAGCGGTCCCGTG

12293

QY

3790

TACTCTGTCATGTGACGCAATATCAGCCCCCGCGCTGCTGAGGACCCATGGCGGCC

3849

DB

12292

CCGCGCGCCCGGACCTCCCGGCGCGCGCTCACCCGGGTTGACCCGACCGCTTCGCGCGGG

12233

QY

3850

GCTCGGCGCTTTCGAATCAGCGGTACAGCGTACCTGTGACCGCGTACAGCGGGGCGGAG

3909

DB

12232

CGGCGCTGCGCGGGGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG

12173

QY

3910

TGGCGCGCTTTCGAGCGCTGGGCGCGGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCG

3969

DB

12172

AGGTCTCGCGCGCGCGCTGTCGCGCGTCCCGTCCGACGAGCGACGTCGCGCTCGGGA

12113

QY

3970

AGGGGTGAGGGGTGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG

4029

DB

12112

CGCACGAGGCGCGCGCGCGCGCGCTCCCGCTCCCGCTCCCGCGCGCGCGCGCGCGCG

12053

QY

4030

GGCGTGCAGCGCGCACCACTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG

4089

DB

12052

ACGCGCTCCCGCGAGGCG

11993

QY

4090

CCGTCGCGCGCGCTGAGCTGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG

4149

DB

11992

GGGCGAGCAGCAGCAGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG

11933

QY

4150

GGGCGAAGGCGGGGAGTCCG

4209

DB

11932

TCTTCG

11873

QY

4210

GC 4211

DB

11872

GC 11871

RESULT 11

AY034378

16643 bp

DNA

linear

BCT 25-MAR-2003

LOCUS

Streptomyces sp. PGA64

silent angucycline-type gene cluster,

partial sequence.

ACCESSION

AY034378

VERSION

AY034378.1

GI:14280338

KEYWORDS

Streptomyces sp. PGA64

SOURCE

Streptomyces sp. PGA64

ORGANISM

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE

1 (bases 1 to 16643)

AUTHORS

Metsa-Ketela M., Palmu K., Kunnari, T., Ylihonko, K. and Mantsala, P.

TITLE

Engineering Anthracycline Biosynthesis toward Angucyclines

JOURNAL

Antimicrob. Agents Chemother. 47 (4), 1291-1296 (2003)

PUBMED

12654660

REFERENCE

2 (bases 1 to 16643)

AUTHORS

Metsa-Ketela M., Kantola, J. and Ylihonko, K.

TITLE

Cloning and Characterization of a Silent Angucycline-type Gene

JOURNAL

Cluster from a Rubromycin B Producing Streptomyces sp. PGA64

Unpublished

REFERENCE

3 (bases 1 to 16643)

AUTHORS

Metsa-Ketela M., Kantola, J. and Ylihonko, K.

TITLE

Direct Submission

JOURNAL

Submitted (11-MAY-2001) Molecular and Cell Biology, Galilaeus Oy, Kairiskulmantie 10, Kaarina FIN-20781, Finland

FEATURES

Location/Qualifiers

1..16643

/organism="Streptomyces sp. PGA64"

/mol\_type="genomic DNA"

/strain="PGA64"

/db\_xref="taxon:161235"

complement(<1..751)

/gene="pgak"

complement(<1..751)

/gene="pgak"

/note="similar to ORF12 from the tylosin biosynthesis

```
cluster"
/codon_start=1
/transl_table=11
/product="Pgaf"
/protein_id="AAK57521.1"
/db_xref="GI:14280339"
/translation="MSGGVAVNQKIRTMATKAGTCAGACRRREAAQOARLAVS
GTVCLPCXERLAREQLPALFECARRLDSRDRVSGGSLPGLPLNTAAVEA
RSATVSLASAAVAQGVAAFTRVVARGFLRLHETLWLAHAHAAGSELSEVAQL
ARGAARVVGPGFNRRLPIGCGVEDCTGRLTATVRPDQPAVPAETICDVSABHRWLGH
QWVLSSRLRGARRDDPAGAGDGAABAEATSAD"
1125..2600
/gene="pgaf"
/feature="CDS"
1125..2600
/gene="pgaf"
/codon_start=1
/transl_table=11
/product="polyketide oxygenase Pgae"
/protein_id="AAK57522.1"
/db_xref="GI:14280340"
/translation="MDAAVIVVAGPAGMMLAGELRLAGVEVVVLERLVERTGESRGL
GFTARTMVFQDGIPLPRFGEVETSTQGHFGLPTDFGLGAWQAATVPSVTETH
LEOWATGLGADIRRGHVLSDTDDGAGVTEVRGEGKHTLRAAYLVGDCGSRVRK
RAGDFPGCTAMTEYLADIKGVELQPMIGETLPGCMVMGCLPGGITRIIVCERGT
PPQRETSPSWEHADAKWRLTGDDIAHAEPWWSAFGNATQVTEYRGRVILAGDS
AHILHPLAGCGQNMNTSIQDANVLGKLGAVNGTATELLDSTHSERHAGVKRLMLNTQ
AQGLFLSGPEVQLRDVLTETIQTGEVARHLGAVSGLEITDYDGTGSHPLLGRMP
ALELTTATRETSSTELLHTARGVLDDADNPLRARAASDRVDIVTAVPGEVSATS
GLRDTTAVLIRPDGHVAAAPGSHHDLPMALERWFGALPG"
2735..3394
/gene="pgaf"
/feature="CDS"
2735..3394
/gene="pgaf"
/codon_start=1
/transl_table=11
/product="putative repressor Pgay"
/protein_id="AAK57523.1"
/db_xref="GI:14280341"
/translation="MGTPQATITANGKAGPAERQDVVRAARSVFAREGWTHTTIDAVA
GBAGTGPDTGVGHFKDQELLSDSSAAALTVVAEDHLGEITDLAODLAAQ
RWMLPLSPFAHFHAIIVSHISAEARLPAGVLEWQNAQPRQAORRELRGLGADQG
LUDIDADAGAAAGRFQLVSAPVVQSRFHGALPLPDEFDTALLADGVADFIRLPSAPA
R"
3462..3791
/gene="pgaf"
/feature="CDS"
3462..3791
/gene="pgaf"
/codon_start=1
/transl_table=11
/product="polyketide cyclase Pgae"
/protein_id="AAK57524.1"
/db_xref="GI:14280342"
/translation="MHSTLIVARMVVESSAQVAELFGDPDRTEMHRMGTRRRQLFSY
NGLYFHLQDFDGDNGELIBEAKSDSRFIRISEDLPKPYEAYDPATWRSPADMAKRF
YDWTATQ"
3829..5097
/gene="pgaa"
/feature="CDS"
3829..5097
/gene="pgaa"
/feature="minimal PKS"
/codon_start=1
/transl_table=11
/product="ketosynthase Pgae"
/protein_id="AAK57525.1"
/db_xref="GI:14280343"
/translation="MSRRVITGVGLAPGGIGAKNFWLSLSEGRATNGITFFDPSS
FRSVAEADDFDEKHGTFQPIRRMDRAQFVAVTAREALDGLDLAGFPHRTGV
TTGSAVGAATGLDDEYRVVSDGRLDVTHTVAPHLVNYFVPSFSFAEVAWAGAEV
PATVVTGCTSGLDVGHATELIRECTADVMITGADAPISPIWACPDATKATPRN
DAAHARSFDETRNGFVLGEGSAFVPLELESACKRGHAIYAEIAGVASCNAFHT
GLRPGDGRMEADVALGARNPNKIDITINAHGSGTKQNDRHETAAFKLSLDHAYR
TPVSSIKSVGHSLGASIEIAASVLMKNHVPTANLHTPDPECDDLDYVPLTARE
QRTDTVLITVSGFGGFSQAMVLARPERNAA"
5094..6311
/gene="pgab"
/feature="CDS"
5094..6311
/gene="pgab"
/feature="minimal PKS"
/codon_start=1
/transl_table=11
/product="chain length determinant PgaB"
/protein_id="AAK57526.1"
/db_xref="GI:14280344"
/translation="MSTRVITGIVATPENGVLGVDFWAAATRVGNKAIARVTHFPDSS
YPARLAGEIRGEAKDHLPSRLIPQTRMTQALVAADCAFEDAGVELGNIIPAYDMGV
VTATSGGFEFGQNELKKLWSQGRVSVAYOSFAWYAVNSQOIGIRNGMGSPGVVV
SDHAGGLDAIAQARQIRKSKLIFSGGFDAASICPWGAAQIAGRLSHVDRPERAYL
FDRRAAGVFPGEAGLILDEBTARERGAEKIYGEVAGYSTFDPGPRGREGCLR
HAIEVALADAGVTAADVDFADAGSPDLDRQADAITAVFGPSGVPVTAFTWIGR
LYSGAAPVDVVAATAVLAIREGLIPTTNVLESPDYDLDLVTGQPRTASVRTLVLARGV
GGFNSAVVRAVD"
6412..6681
/gene="pgac"
/feature="CDS"
6412..6681
/gene="pgac"
/codon_start=1
/transl_table=11
/product="acyl carrier protein PgaC"
/protein_id="AAK57527.1"
/db_xref="GI:14280345"
/translation="MSNNEFTIEDLKRILLEGAGABEGVDLHGDILDTDFEELGVESL
ALLETGGRIERBYGISLDDDFVTQNRTPRTLVDAINGCLKELISS"
6785..7570
/gene="pgad"
/feature="CDS"
6785..7570
/gene="pgad"
/codon_start=1
/transl_table=11
/product="polyketide ketoreductase PgaD"
/protein_id="AAK57528.1"
/db_xref="GI:14280346"
/translation="MSQDQKRVALLVTGTSIGLAVARLLATQNHQVFTGARNAEVVA
ETVQSGDEGLDVGSTLDDVSTDDARAFVQAAVDRFGTVLVNNAGSGGGVTADI
TDLWSDVIDTNLNSVFRILTGLRHKRGRIINIASTAGKQGVVLGAPYSAS
KHGVGFTKALGNELAPGTITWAVCPGVETPMAQRVQGVAAAYDAEAEAILKFKQ
AKIPLGRYSTPEEVAGLVYGLASDTAASTTSQALNVCCGLGNP"
7647..8591
/gene="pgal"
/feature="CDS"
7647..8591
/gene="pgal"
/codon_start=1
/transl_table=11
/product="putative bifunctional aromatase Pgal"
/protein_id="AAK57529.1"
/db_xref="GI:14280347"
/translation="MSQSGLREVEHEITVAAPAAAAYRLIAEWNPRIFPPTTVVQR
EAQCENTRIHWATANGEAKTWSRTTLDPERLITTRQEVSAAPPVAAAGGTWIEP
ASDTSRVRLLHDHYRATDDDDPAKAWIDEADVDRNSRSELAALKENVAHAHAEDLTFE
FEDTIRVNSAKOVDFVYNAQAWPRLPHVATVRFPEPSPGLQLEMDTRAKDGTSH
TTKSYRVTFPHERIAYKQVLTALMTLHTGYTFAEDDAGVAAASQHTVTITANVAR
ILGADATVEDAKKVVQGLSTNSRATLGHAKDYAEARA"
8593..10947
/gene="pgam"
/feature="CDS"
8593..10947
/gene="pgam"
/feature="similar to UrdM from the urdamycin biosynthetic
pathway; contains an additional internal start codon
between the oxygenase and reductase domains"
/codon_start=1
/transl_table=11
/product="oxygenase-reductase PgaM"
/protein_id="AAK57530.1"
/db_xref="GI:14280348"
/translation="MSEODHTEQVVVVVVGAGPVGLFLAGRLRAGADVVLLEOLTAPTPT
ESRSTILHARTWEILDRLGLLASLGDPPNPNMGHFGGIPDLTLPSAYAGQWKVQOTR
IBELLAQWAKDGLADIRRGHEVTALTATGTYAABAEVTAEGRTLELFRACKVCVCGDGEN
```

SVVRLRGFGPPTGEGRELLRADVAGIDVNNRRFERLDGGLAI VARRPPGVTRVMVH	
EPGAVPRGGPEFTDVDAKRVTDGDISGTFPLWNSFDASQADSVDHDFEFLAG	
DAAHQMPIGGALLUGLQDAVNLGWLAAQVAGRGPDGLDSTHGRHVGQOVLN	
IRTQALLGSAEVEPVQVYLAELIANSNVSHLAGAVSLGDIRYDVGPGEHLLGR	
RLPHWQLTDDGAVTSTAALRPGRVGLFLFPGDDPRQARHBEILSAAPAFADRVTV	
TAGSCTGGTDPGTGALLVRPDGVVWADTFPTDLEQALRRWFVFAPTAAANAATPAPAA	
PSTAARANNSSATATAKKWRTHMNRITSKTALVTGSRGMGRAAAIRLAREGALVAV	
HYTKARDAEDVVAIEIKDGRAFTVRAELVPGDVHVELFLGLENLREPTGSDLDI	
LVNNAGVMGVKPEDTTPEKDELFAVNAKAPFFLIORALKNMPDGGRIINISGLTR	
VANPDEIAYMTKGAVDQALAHFAKYLGPNNITVNSVAPGITRNDNVPFPIPAVEOM	
AQLSAFNRVGEPEADVAVFLATDEARWITGSPVDATGGTLLG"	
gene	11068..12534
CDS	/gene="pgaj"
	/gene="pgaj"
	/codon_start=1
	/transl_table=11
	/product="putative transporter Pgaj"
Query Match 7.4%; Score 322.6; DB 1; Length 16643;	
Best Local Similarity 56.0%; Pred. No. 1.7e-19;	
Matches 679; Conservative 0; Mismatches 519; Indels 15; Gaps 3;	
Qy	2964 GCCCCTACGGCCCGCGGGCCCGCTAGCTTGCCTCGCTACCGAAGCGTGCGGG 3023
Db	11004 GCGCCGCGCCGTTACCGACGACCGGGGACGTGGCGTGGTGCTGAGGGGCGGTCA 10945
Qy	3024 TCAGAGAACTCTCCGCTGTGGCATCGACGTACTTGGCCGCTGATCCACCGTGAGTCGTC 3083
Db	10944 GCCGAGCGGTGCTCGGTGCGTTCGACGACGAGGCGGTGATCCAGGGGCGCTCGTC 10885
Qy	3084 GGAGGCCAGAAAGCCACCGTCGGCGATGTCGTGGGTCTGCGGATGCGGTGTAACAC 3143
Db	10884 GGTGCGAGGAAGCCACCGTCCGCGACCTCTCGGTTTCGCGGACGCGGTTGAAGC 10825
Qy	3144 GGAGTTGCGCGCAGTCGCGGGCGCTCGGGGCTCGCGCGCGCTGCTCATGTC 3203
Db	10824 GGACAGTGGCCCATCTGCTCACCGCTCGGGAT-----GCGAAGCGGGGTGTC 10771
Qy	3204 CGTCTCACAGAAACCGCGCGCACCGGCTTGACCGTGATCCCGCTTCCCGCAGTGCCT 3263
Db	10770 GTTGGGTTGATGCGGGCGCACCGAGTTGACCGTGATTTCCGGGGCCCGAGTACTT 10711
Qy	3264 GGCAGGGCGAGCGTGAGCGTGTCACCGCACCTTGTTGTCATCGGTATCCGATGGAATC 3323
Db	10710 CGCGAAGTGCAGGGCGAGCTGTCGACGCGCCCTTGCTGTCATGGGTACGGCATCTCGTC 10651
Qy	3324 GGGGAACGCGCGCGGTGCGGGCAGACAGATGTTGATGATCGCCCGCGCTCGCGAG 3383
Db	10650 CGGGTTGCGACCCGGGTGAGCGCGGAGTGATGTTGATGATGCGCCCGCTCGGGCAT 10591
Qy	3384 TCGTTTCAGTCCGTCTCGACACAGAAACAGCGGTGCCCGGACGTTGACGGGACAGTCG 3443
Db	10590 GTTCTTAGCGCGGCTGATGAGAGAGAACCGCGCCTTGGGTTGACCGGAGAGTTC 10531
Qy	3444 GTCGAAGACCTCTCTCGTGATCTTCGTTGATCCGTCGAGCGCGTGAAGCCCGCTTGT 3503
Db	10530 GTCGAACCTCTCTCGGGTGGTGTCTCTCGGGCTTGAACGCGGCCCATGACACCGCGCTTGT 10471
Qy	3504 CACGAGATGTCGACTCGGGGGGACCTCCGAACCTCGCCCATCCCGGGGTGMAACGCGC 3563
Db	10470 GACGAGGATGTCAGGTGCGGTGAGCCGGTGGCTCGCGAGCGCTTCTCAGGCCGAG 10411
Qy	3564 GTAGAGCGCGCGCGGTCAACCCACAGCCGAGTTTCGGCCGGATGGGCAACGCTGTCC 3623
Db	10410 GAAAGATTCTGTCAGCTCGCCCGGAGCGCGAGTTCCGGGGTACGGTGAAGGACGCGC 10351
Qy	3624 GCCGTGCTCCGGATGTTGTCGACGGTCTCTCGCGCGCGCGCTCGCTGCTGCCGTAGTG 3683
Db	10350 ACCGTCTTCTCGATCTCGGCGACGACGCTCTCTCGCGCGCGCTCTTTCGCGCGGTGTAGTG 10291
Qy	3684 GACTGCCAGGAGCGCCCGTCCGCGCGCAGCGGAGGAGTACCGGTCCGATGCCCGC 3743

Db	10290 GACCGCGACGAGTGGCGCCCTCCCGGCGAGGCGTATCGCGCGACCCCGGCCATGCCCT 10231
Qy	3744 GCTTCCCGCGGTACACAGGGCGGTCTTCCCTCCAGCGGTCTTCCATACCTCGTCCCATG 3803
Db	10230 GCTGAGACCGGTGACACGCGCGGTCTTCTGTCGTGAGCGCTTTCATGTGTCTTCCACTT 10171
Qy	3804 TGACGCGATATCAGCCCCCGCTGGGTGAGCGACCCATGCGCGCGCTCGGCGGTTCGA 3863
Db	10170 CTTCCGCGTCCGTGTGGCCGACGAGTTGTTGTCGCCCTCGC-----CGCGGTGACGGT 10116
Qy	3864 ATGACCGGTACAGCTTACCTGTGACCGCGTACAGCGGGCGGAGTGCGCGGTTCGAC 3923
Db	10115 GCGCGCGGGCGGCGCGGTGGCTGCGCGCGCGCGCGCTTTCGAAACACGACGCGCGC 10056
Qy	3924 GCGTGGGGCCAGATCGGGCGCGCGCACGGGGAACCGCGCGCGTCAAGGGTCAAGGGGTC 3983
Db	10055 AGC-----GCCTGTTCCAGGTCCGTGCGCGGGGTGTGCGGCCAGACGATATCCGTGGG 10000
Qy	3984 GCGCGGACCCCGCGGTTCAGGGCACCGATCGAGTTCGAGTTCGGCGGTGCCACGCG 4043
Db	9999 CCGGACGAGGCGCGCGGTGTCGGGTGTCGGCGCGCGGTGACAGCGCGGTGAC 9940
Qy	4044 CCACCACTCTCGCGGCCAGCTCCGACTCGTACCGGTACAGAGCCGCTCCGCGCGGAG 4103
Db	9939 GGTGCGCACCGGTCCGCGCGCGGGCGCGCGGCGGAGCGAGTTCCTCTCGTCCGCGGCTG 9880
Qy	4104 TCTGAGCTGACGTCGCGCGCGCGGTGGGTGAGCGGTTCGCGCGCGGCGGAGGCGGG 4163
Db	9879 CTCGCGCGGTGCTGCGCGGGCGACGAGGAGAGACCGCGCGCGGGGCGCGGCGCGC 9820
Qy	4164 GAGTCCGCGCGCG 4176
Db	9819 CGTCTGCTGACG 9807
RESULT 12	
AF322256/c 39428 bp DNA linear BCT 29-OCT-2002	
LOCUS Streptomyces antibioticus simocyclinone biosynthetic gene cluster,	
partial sequence.	
AF322256	
AF322256.1 GI:16223998	
Streptomyces antibioticus	
Streptomyces antibioticus	
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
Streptomycineae; Streptomycetaceae; Streptomyces.	
REFERENCE	
1 (bases 1 to 39428)	
Galm, U., Schimana, J., Fiedler, H.P., Schmidt, J., Li, S.M. and	
Heide, L.	
Cloning and analysis of the simocyclinone biosynthetic gene cluster	
of Streptomyces antibioticus Tu 6040	
Arch. Microbiol. 178 (2), 102-114 (2002)	
REFERENCE	
2 (bases 1 to 39428)	
Galm, U., Li, S.-M., Schimana, J., Fiedler, H.-P. and Heide, L.	
Direct Submission	
Submitted (17-NOV-2000) Pharmazeutische Biologie, Pharmazeutisches	
Institut, Mikrobiologie/Biotechnologie, Universitaet Tuebingen, Auf	
der Morgenstelle 8, Tuebingen 72076, Germany	
Location/Qualifiers	
1..39428	
/organism="Streptomyces antibioticus"	
/mol_type="genomic DNA"	
/strain="Tu 6040"	
/db_xref="taxon:1890"	
<1..168	
/gene="sim2"	
<1..168	
/gene="sim2"	
/note="ketosynthase"	
/codon_start=1	
/transl_table=11	
FEATURES	
source	
gene	
CDS	

```
/product="Sim2"
/protein_id="AAL15580.1"
/db_xref="GI:1622399"
/translation="IPPTANLHDPEDCDLVVPTITAREARLDRVLVSGSGFGFQSA
WVLRPERNTA"
165..1379
/gene="sim3"
CDS
165..1379
/notes="chain length determinant"
/codon_start=1
/transl_table=11
/product="Sim3"
/protein_id="AAL15581.1"
/db_xref="GI:1622400"
/translation="MTTSVVTGLGVAAPNGLGTADYWAATREGRSIGRVRTFPDPSQ
YPSRAGVPGFVAEDHLPSRLLPDTHMTRIALVSADWLQDAGIRPEELPEVAAGV
VTASAGGFPGQNELRALWSKGSQYVSAYQSFAWYAVNTGQISIRHGLRPGSGVV
SDQAGLDALAAQRRQIRGSGQIVSGGIDASICSWGGAQTSGRMSTDDDPATAYR
PFDGAAGHVPCEGALVLEAGAAAAGAVYGRYAGYGTTFDPKSGSGREPLRR
AIELATDAGLAPADIVVFADAAVPLDRIEARALVEVFGPLPYTAPKSMTGRL
NSGAGSLDVATALLAIRDGVPIPTINVTQDDYELDLVTAQARSARLSALVARGHG
GFNSALVQDAA"
1428..1694
/gene="sim4"
CDS
1428..1694
/notes="acyl carrier protein"
/codon_start=1
/transl_table=11
/product="Sim4"
/protein_id="AAL15582.1"
/db_xref="GI:1622401"
/translation="MQTTEPTLDDLRRLRDGAGASEVDVLEGGDIIDVAFEDLGYESI
ALLETCSRIRERYGISLDDDAVSEAPTPRALVAIVNTHLIAVAG"
1714..2499
/gene="sim5"
CDS
1714..2499
/notes="ketoreductase"
/codon_start=1
/transl_table=11
/product="Sim5"
/protein_id="AAL15583.1"
/db_xref="GI:1622402"
/translation="NPGQQQVALVTGATSGIGLAVRALAAQGHQVFCARNAKVA
VTIKELRGGLVDGTSCDVSVDKAFVAAAVRPFVVDVNNAGSGGVTAEAL
TDELWADVIETLNTSVFMVTRVLTAGGRKSSGRIVNIASTAGKQGVVLGAPYAS
KHGVVGTALGNELAPGTITVNVCPGYVETPMABRVQGYAAAYGTSEAILBKFO
AKIPLGRYSTPEBAGLVGLVSDTAASITAQALNVCGGLGNF"
2525..3472
/gene="sim6"
CDS
2525..3472
/notes="cyclase/dehydrase"
/codon_start=1
/transl_table=11
/product="Sim6"
/protein_id="AAL15584.1"
/db_xref="GI:1622403"
/translation="MSDGSISHVEHTIEVDAPATDVIYRLLAEVWNPRIFGPTIIVHDH
VERGARTERIRIWTANGAAKSWTSRELDPEALRIDFQEVSAAPPVAAAMGAWIIEP
VSEGACRVLHLDHYRAIDAPEDLAWIEQAVDRNSRSELAALKANAEVVVGADELML
SFQDSVRIDGSAKDVIYQTVLPALMTLHTGYLLEJDEGEGTVATSQHTVIINTANIR
HTTSSVRVCFPHRIYKQTVLPALMTLHTGYLLEJDEGEGTVATSQHTVIINTANIR
AILGSGAGVAEAEARQYVQTALSTNSRATLGHAKDYAERD"
3478..4962
/gene="sim7"
CDS
3478..4962
/notes="oxygenase"
/codon_start=1
/transl_table=11
```

```
/product="Sim7"
/protein_id="AAL15585.1"
/db_xref="GI:1622404"
/translation="MTGPTDVLVIGAGPVGLMLAGELLGLGARVVVAERLTETPTTES
RASTLHARTMEILQDQRLTGLGEVPGDGTGHFGGIPLRDQQQSPYAGLMKVPQSRV
EILGAWAIGLGAFAVRGRELTLGRDGAUVVEAELSCDGTRELLRARYVVGCDGSENS
TVRLAGPEFAGSAASRELIRADVAGIDVNNRRPERLPGLLAIARRPDGITRVWVHE
FGRTATPQAPDPDFDEMAQTWQRTVGEDISHTGTPWINSFGDASQSTVEYRGRILLA
DGAHAQMPVGGQALNGLQDAFNGLKLAABICGNWAPPGLDLSYHEERHAGSVRLS
DIAAQAALLGGPEVALRSVGRLLGLDADVDHLATGVGGIGIYSAHBGHPLLIGA
RLPDAALDIEAGAATAWALLRAGTGCLLDVAPEAPGHVDLRAVADGWADRVRVAARP
RAGAVATPTGAVLVRPDGHVVWADGGADRLRAALVQWFGSADRT"
5041..5805
/gene="sim8"
CDS
5041..5805
/notes="oxidoreductase"
/codon_start=1
/transl_table=11
/product="Sim8"
/protein_id="AAL15586.1"
/db_xref="GI:1622405"
/translation="MKRLAGKTALVTGSSRGIGRATRLRLAREGATVAVHYADNADAA
KATLAAIEEGRAFLVRAELGVPDVTFLQGLTAGREATGKEELNILVNNAGVMG
GVKPEDLTPEIFDRLIAVNAKAPLFTQRLVLSIMPDGRIINIGTGLLRANPDEVAY
SMSKALDQISILHFAILLPRNITVNTVSPGITTNGSPIFDVPEAREFMANNAFERRI
ABPDVDGVAFPLASDADARWITGTFTVDASGGTLLGG"
5936..6508
/gene="sim9"
CDS
5936..6508
/notes="reductase"
/codon_start=1
/transl_table=11
/product="Sim9"
/protein_id="AAL15587.1"
/db_xref="GI:1622406"
/translation="MSSARLRLAVLVGSTREGRLCPTVANWFAGQGRHQGGFDLLDLD
PYGLDIPVAPSPPEPQVVEIGQRLAADAFVITVPEYNHSPSAKLILIDHYVOMR
AKPVGFSYGGGLGRLAVEQLRVFAELHAATVRDVTVSFHNAQLFDEAGDPRDPEG
CNTAAKTMLDOLHWMGRALRQLRERYQG"
6516..7202
/gene="sim10"
CDS
6516..7202
/notes="putative phosphopantetheinyl transferase"
/codon_start=1
/transl_table=11
/product="Sim10"
/protein_id="AAL15588.1"
/db_xref="GI:1622407"
/translation="MARTESVQPAEAVQLWLRPFPVRPDELAALDSEERRRABAFPG
SADRMLYLSAHLARLRLAARLHVAPQEVIRVDRNGRPTLPGRPPPHFSLSHSAG
LALLCTAPVRIQVDVQRTLSRTTADLCGRHLHPASOELASVQPSARAAHPTRLWTRK
EAYLKGUGVGLSRSLAADYLVGRDPTPAGTIVLDVPCGESHLASAAVQSRSLPVTLL
RWLLRSADHV"
7195..8781
/gene="sim11"
CDS
7195..8781
/notes="carboxyl transferase/decarboxylase"
/codon_start=1
/transl_table=11
/product="Sim11"
/protein_id="AAL15589.1"
/db_xref="GI:1622408"
/translation="MSDVPSPADPAPTRCRLEELRALRQAVADGDPDPRAEAQAHK
GKLTARERVELLDGSGFHEVRLRHRQAQGFGLKARKPHTDGVVVGWTVHGRIVFV
YAHDPRIFGALGEAQAQIKHIMDAIAGAPLVSLNDGAGARIOGTPTALAGYGI
FORNTRASGVIPOISVMLGPCAGGAAYSPALDVFVWRETSQMFITGDPVQVATGE
KLSQNLGGADVHSAGSVSHFVDDQSCIEBVRVLLSLLPSNNREMPVQLSDDPV
DRNDSLDDLVDPADGNRPVIEEVVDYGEYIEVHERWATNVLCLARIDGHVTG
IIRANQPSLAGLIDISASEKAARFVQMCDAFNILPLVTLDDVPGFLPFGVQSHGGIIRH
```



GAKLLYACNATVPRIISILIRKAYCGAYIVMDSRSISIGADLAYAWPTNIEIAVMGEGAA  
 NVIFRRDIAAEDPDAMRTOKIKEYKAEIMPHYAAERGLVDDVIDPAETRQVLTSAL  
 AMLRTKHADLPSRKHGPPM"  
 8796..9014  
 /gene="sim12"  
 8796..9014  
 /gene="sim12"  
 /note="hypothetical 7.6 kDa protein"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="Sim12"  
 /protein\_id="AAL15590.1"  
 /db\_xref="GI:16224009"  
 /translation="MTSDAMTDATRVESGHLEPELAALTVVLLARIRGAAPDRAAGD  
 GGSIVAAIWRRLRANGQSPSWRTAC"  
 9056..9829  
 /gene="sim13"

Query Match 7.3%; Score 316; DB 1; Length 39428;  
 Best Local Similarity 62.1%; Pred. No. 4.6e-19;  
 Matches 518; Conservative 0; Mismatches 310; Indels 6; Gaps 1;

QY 2977 CCGGCGGCGGCGCCGTAGTCTGCCCTCGTACCGAAGCGTGGCGGGTFCAGAGAATCGTT 3036  
 |||||  
 DB 5846 CAGGTGCCGCGCGCGTCCGATCGCGCGGCGCGAGCGGTCAGCGCCCGCAGCGGTG 5787  
 |||||  
 QY 3037 CCGCTGTGGCATACAGTACTGGCGGTGATCCACCTGAGTGTGCGAGCGCAGAAAG 3096  
 |||||  
 DB 5786 CCGCGCTGGCGTCCAGAAAGTGCCTGTATCCAGCGGGCGTGTGCGAGCGCAGGAAG 5727  
 |||||  
 QY 3097 GCCACAGTGGCGGATCTGTCGGGCTCGCGGATCGGTTGAACACGAGTTCGCGGCC 3156  
 |||||  
 DB 5726 GCCACAGTCCCGAGCTCTCGGGCTCGGGATCCGCTTGAAGCGTTCACTGGGCC 5667  
 |||||  
 QY 3157 AGTGGCGGCGGCGCTCGGGGCTCTGCGCGCGCGTGTGCTCATGTCGTCTCCACGAAA 3216  
 |||||  
 DB 5666 ATGAATCCCGCGCTCGGGAGCTCGAAGATCGGGTCCGCTGTGTTGGT-----GATG 5613  
 |||||  
 QY 3217 CCGGCGGCGACCGGTTGACCGTGTATCCCGCTTCCCGAGTTCGTCGCGCAGCGAGC 3276  
 |||||  
 DB 5612 CCGGCGGACACCGTGTGACCGTGTATGTCGCGGCGCGAGCAGCGGCGGAGTGCAGG 5553  
 |||||  
 QY 3277 GTGAGCGTGTCCACCGCACCTTGTGTATCGGTATCGGTATCGGTGACTCGGGGAAACGCGCG 3336  
 |||||  
 DB 5552 CTGATCTGGTCGAGCGCGCTTGTCTCATCGATGAGGCGACTCTGTCGGGGTTCGCGTAA 5493  
 |||||  
 QY 3337 CCGGTTCGCGGCGAGACAGATGTTGATGATCGCGCGCGTTCGCGAGTCTGTTTCAGTCCG 3396  
 |||||  
 DB 5492 CGCAGCAGTCCGTACCGATGTTGATGATACGTCGCGCGTTCGCGCATGATCGACAGCAG 5433  
 |||||  
 QY 3397 TGTGGAACAAGACAGCGGTGCGCGAGCGTGTGACGCGGACCGACGTCGTCGAAGACCTCC 3456  
 |||||  
 DB 5432 CGCTGGGTTCAGAAAGCGGAGCTTGGCGTTGACCGCGATGAGCGCGTTCGAAGATCTCC 5373  
 |||||  
 QY 3457 TCGGTGACTTCGTATCGTCCGAGCGGCTGACGCGCGGTGTGTTTACCAGGATGTCG 3516  
 |||||  
 DB 5372 GGGGTGAGGTCTCTCGGCTTCACGCGCGCCATCACACCGCGGTGTTTACCAGGATGTCG 5313  
 |||||  
 QY 3517 AACTCGGCGGCGACATCCGAACTCGCCCATCCCGGCGTTCGAAACCGCGCTAGAGCGCGCC 3576  
 |||||  
 DB 5312 AGCTCTCTTCCCGTTCGCGCTCGCGACACTCGCGTCAGACCCCTGGAAGAGGCGTTCG 5253  
 |||||  
 QY 3577 GGTACCCACGACCGCGAGTTCGCGCGCGATGCGCAACCGCTGTCGCGCTGCTCCGG 3636  
 |||||  
 DB 5252 AGTACCGCGGACACCCAGCTCGCGCGCGACAGGAAGCGCGCGCCCTCTTCTTCCTCG 5193  
 |||||  
 QY 3637 ATGGTCTGACGGTCTCTCGCGCGCGCTCGCTGCTGCGTGTGCGTGTGCGTGTGCGGCG 3696  
 |||||  
 DB 5192 ATCGGCGCAATGTCCTTCTGCGCGCGTGTGCGGCTGTGCGGCTGTGCGGCGGCGACATC 5133  
 |||||  
 QY 3697 GCCCGCTCGCGCGCGCGGCGGATACCGCGTTCGATGCCCGCGCTTCCCGCGTCC 3756  
 |||||

Db 5132 GCGCCCTCGCGCGCCAGCGCGCGGTGCGCCCGACCGATACCGCGGCTTGATCCGGTC 5073

QY 3757 ACCAGGCGGTCTTCCCTCCAGCGGTCTTCCATACCTCGTCCCATGTGCAAGC 3810  
 |||||  
 Db 5072 ACCAGGCGGTCTTCCAGCGCGGTGCGCTTACAGAGTCTCTCTCGTCCGATGC 5019  
 |||||

RESULT 13  
 AF324838/c 34869 bp DNA linear BCT 14-NOV-2003  
 LOCUS Streptomyces antibioticus simocyclinone biosynthetic gene cluster,  
 DEFINITION partial sequence.  
 ACCESSION AF324838  
 VERSION AF324838.1 GI:12744817  
 KEYWORDS Streptomyces antibioticus  
 SOURCE Streptomyces antibioticus  
 ORGANISM Streptomyces antibioticus  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Streptomycineae; Streptomycetaceae; Streptomyces.  
 1 (bases 1 to 34869)  
 Trefzer, A., Pelzer, S., Schimana, J., Stockert, S., Bihlmaier, C.,  
 Fiedler, H.P., Welzel, K., Vente, A. and Bechtold, A.,  
 Biosynthetic gene cluster of simocyclinone, a natural multihybrid  
 antibiotic  
 Antimicrob. Agents Chemother. 46 (5), 1174-1182 (2002)  
 11959542  
 2 (bases 1 to 34869)  
 Trefzer, A. and Bechtold, A.,  
 The simocyclinone biosynthetic gene cluster isolated from  
 Streptomyces antibioticus Tue6040  
 Unpublished  
 3 (bases 1 to 34869)  
 Trefzer, A. and Bechtold, A.,  
 Direct Submission  
 Submitted (29-NOV-2000) Pharmazeutische Biologie,  
 Christian-Albrechts-Universitaet Kiel, Gutenbergstr. 78, Kiel  
 24118, Germany

FEATURES  
 Location/Qualifiers  
 1..34869  
 /organism="Streptomyces antibioticus"  
 /mol\_type="genomic DNA"  
 /strain="Tue6040"  
 /db\_xref="taxon:1890"  
 /clone="cosmid 5J10"  
 1337..2812  
 /gene="simA7"  
 1337..2812  
 /gene="simA7"  
 /note="involved in polyketide formation"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative oxygenase SimA7"  
 /protein\_id="AAK06782.1"  
 /db\_xref="GI:12744818"  
 /translation="MDAQIVVVGAGPAGLMLAGELRAGVVDVWVLERLAQPTGSRGL  
 GFTARTLEVDFORGLPRLGDIETDLHGFGGLPLDIFSLIPAHSGGKGVPOAVTEA  
 MLDEWLELGD IRRGHELIISFTDNGDFVEVEVATSDGQRLRTAFIVGCGGSRVTR  
 KAAGDFPDSATMEMFLGVDVRCGDIPTRMIGETVPGGMVWAAPIGNGVDRIIVCEHG  
 TPARRDEPPSEVESAARQLRTQDISAGEPVWLSAFGNPARQVTEYRGRVLLAGD  
 AAHTLHPAGGGMNTSIQDVVNLGKLAATVKGAPGGLLOSIRYDLGPGDHPLLGLRI  
 QAQGLFLSGAEMQFLRDVMSELITFENVSRHLAGMVGSLIRYDLGPGDHPLLGLRI  
 PHLELVGDAKGTSTTELLHPARGVLLDADAGLRETAAPWADRVVDVVVTVTHGAADA  
 POYEPADALLIRPDGHVAAWAPGPVALSEALARSFCAPRSR"  
 2862..3188  
 /gene="simA4"  
 2862..3188  
 /gene="simA4"  
 /note="involved in polyketide formation"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative cyclase SimA4"  
 /protein\_id="AAK06783.1"  
 /db\_xref="GI:12744819"

gene  
 CDS





Db	10071	CGCCGCTGGCGTGCACAAAGTGCCTGATCCAGCGGGCGTGGTTCGAGCGCGAGAG	10012
Qy	3097	GCCACACGTCGGGAGATGTCGCGGTCTGCGATGCGGTTGAACAACGAGATGTCGCGCC	3156
Db	10011	GCCACACGTCGGGAGATGTCGCGGTCTGCGATGCGGTTGAACAACGAGATGTCGCGCC	9952
Qy	3157	AGTGCCTGGCGGCTGCGGCGGTCTGCGCGCGCGCTGCTTCATGTCGCTTCCACGAA	3216
Db	9951	ATGAATCTCCGCGCTTCCGCGGACGTCGAAGATCGCGGCTGCGCTTTCGTTGCT	9898
Qy	3217	CCGCGCGCACCGCTTACCGCTGATCCCGCTTCCCGCAGTTCGCTCGCCAGGCGCAGC	3276
Db	9897	CCGCGCGCACCGCTTACCGCTGATCCCGCTTCCCGCAGTTCGCTCGCCAGGCGCAGC	9838
Qy	3277	GTGACGCTGTCCACCGCACCTTGTGATCGCGTATCCGATGCGGACTCGCGGAAACGCGCG	3336
Db	9837	CTGATCTGCTGCGCGCTTGTGATCGATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG	9778
Qy	3337	CGGTCGCGCGCACGAGATGTCGATGATCCCGCGCGCTGCGCGCTGCGGCTGCGGCTG	3396
Db	9777	CGCAGCAGTCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	9718
Qy	3397	TGCTGGACAGCAACGCGCTGCGCGGCTGACGCGCACGCTGCGCTGCGGACCTCC	3456
Db	9717	CGCTGGCTCAGGACAGCGGAGCTTGGCTTGCACGCGATGAGCTGCTCGAGATCTCC	9658
Qy	3457	TCGCTGACTTCGCTGATCCGTCGCGGCTGACGCGCGCTGCTGCTGCTGCTGCTGCTG	3516
Db	9657	GGGCTGAGTCTCTCGGCTTCAACGCGCGCTTCAACGCGCGCTTCAACGCGCGCTT	9598
Qy	3517	AATCGCGCGGACCTCCGAACTCGGCTTCCGCGCTGCGGCTGCGGCTGCGGCTGCGG	3576
Db	9597	AGCT	9538
Qy	3577	GCTCACCACAGCGCGCTTTCGCGCGGCTTTCGCGCGGCTTTCGCGCGGCTTTCGCG	3636
Db	9537	ACGTACCCCGGACACCCAGCTCGCGCGGCTTTCGCGCGGCTTTCGCGCGGCTTTCG	9478
Qy	3637	ATGCTCTGAGGCTCTCTCGCGCGGCTTTCGCGCGGCTTTCGCGCGGCTTTCGCGCG	3696
Db	9477	ATCGCGCGGCTTCT	9418
Qy	3697	GCCCGCTCGCGCGGCTTTCGCGCGGCTTTCGCGCGGCTTTCGCGCGGCTTTCGCGG	3756
Db	9417	GCGCGCTCTCGCGCGGCTTTCGCGCGGCTTTCGCGCGGCTTTCGCGCGGCTTTCGCG	9358
Qy	3757	ACCAGCGGCTTCT	3810
Db	9357	ACCAGCGGCTTCT	9304

RESULT 14	
LOCUS	BX640414/c
DEFINITION	Bordetella pertussis strain Tohama I, complete genome; segment 4/12.
ACCESSION	BX640414 BX470248
VERSION	BX640414.1 GI:33571793
KEYWORDS	complete genome.
SOURCE	Bordetella pertussis Tohama I
ORGANISM	Bordetella pertussis Tohama I Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Alcaligenaceae; Bordetella.
REFERENCE	1
AUTHORS	Parkhill,J., Sebahia,M., Preston,A., Murphy,L.D., Thomson,N., Harris,D.E., Holden,M.T.G., Church,C.E., Bentley,S.D., Mungall,K.L., Cerdano-Tarraga,A.M., Temple,L., James,K., Harris,B., Quail,M.A., Achtman,M., Atkin,R., Baker,S., Basham,D., Bason,N., Cherevach,I., Chillingworth,T., Collins,M., Cronin,A., Davis,P., Doggett,J., Feltwell,T., Goble,A., Hamlin,N., Hauser,H., Holroyd,S., Jagels,K., Leather,S., Moule,S., Norberczak,H., O'Neill,S., Ormond,D., Price,C., Rabinowitz,E., Rutter,S.,

Sanders,M., Saunders,D., Seeger,K., Sharp,S., Simmonds,M., Stelton,J., Squares,R., Squares,S., Stevens,K., Unwin,L., Whitehead,S., Barrell,B.G. and Maskell,D.J. Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica Unpublished 2 (bases 1 to 343243) Sebahia,M. Direct Submission Submitted (06-AUG-2003) Submitted on behalf of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: ms@sanger.ac.uk Location/Qualifiers 1. 343243 /organism="Bordetella pertussis Tohama I" /mol_type="genomic DNA" /strain="Tohama I" /db_xref="taxon:257313" 194..2230 /gene="tktA" /locus_tag="BP0999" /note="synonym: tkt" 194..2230 /gene="tktA" /locus_tag="BP0999" /EC_number="2.2.1.1" /note="Similar to Escherichia coli transketolase 1 TktA or Tkt or B2935 SW:TKT1_ECOLI (P27302) (663 aa) fasta scores: E(): 5.7e-137, 63.433% id in 670 aa, and to Pasteurella multocida transketolase 1 TktA or Pml242 SW:TKT1_PASMU (P57927) (668 aa) fasta scores: E(): 3e-158, 61.504% id in 678 aa" /codon_start=1 /transl_table=11 /product="transketolase 1" /protein_id="CAE41300.1" /db_xref="GI:33571794" /db_xref="GOA:Q7VZC0" /db_xref="InterPro:IPR005474" /db_xref="InterPro:IPR005475" /db_xref="InterPro:IPR005476" /db_xref="UniProt/TREMBL:Q7VZC0" /translation="MSNPAPKALADAVKANSHPGPMGMAEIAQAL WLGMLRNPDADPANDRPFVLSNGHSMILYALLHLTGVDLPLEELKNFQLHSRTP GHPEVITPGVETITGPGGLANAVGALAEALAEAFNRAGHTVVDHHTYVFLGDS CLMEGI SHEACSLAGTILKSLVALYDDNGISIDGHVEHWFADDTATFEGYGNVIR GVGDHVAADVDAIKQAQSEKPTLIVCRTVIKGSFPMAGTTHVHGAFLGNDEIAA TRAAAGVAGSFPVQDIYAADSRPAGALAAQAWEDKFAQYAAQYPAFAAEFTRMK GELPEGPAARFEALQATVAKGETVATRKASQFAIAELAGLLPEMLGGSADLTGSNFT DWKGAFLRAGHGVHFGRIHNYGVRSFGMAIIMNGIALHGGYLPFGGTFETFSYSR NAIRMAALMKQRVVHVTHDSIGIGEDGPTQHSIEHAASLRLI PNL5VWRPCTVETA AANGAVTRPASIGMDLHDGPTALLSLRQNLFPFARDEATQKAIVRGYYVLRLDAHA RAAIIATGSEVAIALAAQLAADGIAVRVSMFSTDVDFQDAEWEKVEVLPAQMPRV AVEAGVTAFWHKYVLEGAVVGVIDRYGESAPAGALFKFGLTAEKVAEAVKQVL" 227..2203 /gene="tktA" /locus_tag="BP0999" /note="HMPFfam hit to PF00456, Transketolase, thiamine diphosphate binding domain" 239..301 /gene="tktA" /locus_tag="BP0999" /note="ScanKegexp hit to PS00801, Transketolase signature 1. Confirmed by InterPro eMOTIF pattern match." 2248..3258 /gene="gap" /locus_tag="BP1000" /note="synonym: hexC" 2248..3258 /gene="gap" /locus_tag="BP1000" /EC_number="1.2.1.12"
--

```
/note="Similar to Pseudomonas aeruginosa glycerolaldehyde
3-phosphate dehydrogenase Gap or HexC or paf195
SWALL:G3P_PSEAE (SWALL:P27726) (334 aa) fasta scores: E():
9.1e-84, 66.07% id in 336 aa, and to Alcaligenes eutrophus
glycerolaldehyde 3-phosphate dehydrogenase, plasmid cbbgp
SWALL:G3PP_ALCEU (SWALL:P50322) (336 aa) fasta scores:
E(): 1.1e-107, 84.52% id in 336 aa"
/codon_start=1
/transl_table=11
/product="glycerolaldehyde-3-phosphate dehydrogenase"
/protein_id="CAE41301.1"
/db_xref="GI:33571795"
/db_xref="GOA:Q7VZB9"
/db_xref="InterPro:IPR000173"
/db_xref="UniProt:IPR006424"
/db_xref="UniProt/TREMBL:Q7VZB9"
/translation="MTIRVINGYGRIGRNLRAHYHKKHDIIEIVINDIGDPKTN
AHLTFEDTAHGKPGCTVGDGVVNGDKIRVLNRPAPLPMWELGVGVVLECTGF
FTSKKAGNLKGGAKVLIISAPGKDVDAIVFGVNGQNLKAEHTVINSACTNCL
APLVKPLHDKLGVETGLMTTHAYTNDQVLTIDVHEDLRKASATSMIPKTIQAAAL
VGLVPELNGKLDGFAIRVPTINVSILVLSFVAKGDTTVEEVNSILOAAAGELKDIIL
TYNTEPLVSIIDFNHPASSNFATLTIKVSGKLKVKVSSVYDNEWGFSNRMLDTTVALMS
AK"
2251. .3252
/misc_feature
/locus_tag="BP1000"
/note="HMMPfam hit to PF00044, Glycerolaldehyde 3-phosphate
dehydrogenase, NAD binding domain"
2704. .2727
/misc_feature
/locus_tag="BP1000"
/note="ScanRegexp hit to PS00071, Glycerolaldehyde
3-phosphate dehydrogenase active site. Confirmed by
InterPro eMOTIF pattern match."
3376. .4569
/misc_feature
/locus_tag="BP1001"
/note="pgk"
3376. .4569
/misc_feature
/locus_tag="BP1001"
/note="pgk"
3376. .4569
/locus_tag="BP1001"
/EC_number="2.7.2.3"
/note="Similar to Escherichia coli phosphoglycerate kinase
Pgk or B2926 SW:PGK_ECOLI (P11685) (386 aa) fasta scores:
E(): 6.3e-83, 64.042% id in 381 aa, and to Alcaligenes
eutrophus phosphoglycerate kinase, chromosomal cbbkC
SW:PGKC_ALCEU (P50319) (413 aa) fasta scores: E():
1.8e-103, 75.381% id in 394 aa"
/codon_start=1
/transl_table=11
/product="phosphoglycerate kinase"
/protein_id="CAE41302.1"
/db_xref="GI:33571796"
/db_xref="GOA:Q7VZB8"
/db_xref="InterPro:IPR001576"
/db_xref="UniProt:IPR001576"
/db_xref="UniProt/Swiss-Prot:Q7VZB8"
/translation="MSNVNTSALAKSGKRVFIRADLNVPFDDAGRISDTRIR
ASVPGRLDALGAAVMVTVSHLGRPKGALTEADSLAPVQRLSELGQVRLVDPDW
DGVSVRPGVLDAGVNVGSKDDDEGLSRKMAALCDVYVNDAPGTRARAEATTHGI
ARFAPVACGLLEALDNLGRLHDPKRPVIAIVGGSKVSTKSLIQSLADKVDOLV
VGGGTANTPMLAAGPIGKSLAEPSPQVQRAVIEIMKRGARVPIPTDVVCAK3FGA
DAAATVKAADVAEDMILDIGPQTAQRLADILKTAGTIVWNGPVVFEDQFAHGT
VVARAIDASGFSIAGGSDTLAAIAKYGIADQGTGYISTGGGAFLEFLGKALPAVAL
QARA"
3421. .4563
/misc_feature
/locus_tag="BP1001"
/note="pgk"
3433. .3465
/misc_feature
/locus_tag="BP1001"
/note="pgk"
3433. .3465
/locus_tag="BP1001"
/note="ScanRegexp hit to PS00111, Phosphoglycerate kinase
signature. Confirmed by InterPro eMOTIF pattern match."
4066. .4089
/misc_feature
```

```
/gene="pgk"
/locus_tag="BP1001"
/note="ScanRegexp hit to PS00017, ATP/GTP-binding site
motif A (P-loop)."
complement(4599. .5372)
/gene="fabI"
/locus_tag="BP1002"
/note="synonym: envM"
complement(4599. .5372)
/gene="fabI"
/locus_tag="BP1002"
/EC_number="1.3.1.9"
/note="Similar to Escherichia coli, and
enoyl-l-acyl-carrier-protein) reductase [NADH] FabI or EnvM
or B1288 or Z2512 or Ecs1861 SW:FABI_SCOLI (P29132) (261
aa) fasta scores: E(): 1.1e-35, 41.200% id in 250 aa, and
to Rhizobium meliloti putative enoyl-acyl-carrier-protein
reductase NADH TR:CAC41683 (EMBL:AL591783) (268 aa) fasta
scores: E(): 3e-40, 44.578% id in 249 aa"
/codon_start=1
/transl_table=11
/product="enoyl-l-acyl-carrier-protein) reductase [NADH]"
/protein_id="CAE41303.1"
/db_xref="GI:33571797"
/db_xref="GOA:Q7VZB7"
/db_xref="InterPro:IPR002198"
/db_xref="UniProt/TREMBL:Q7VZB7"
/translation="MNAAPLQGRRLVGTIGIANQDSIANGCARAFALGAELAVTYLN
DKARVHVPLARSIDAAALLPLDLRLVDGQLSFVEQIGARMGSLDFVHMSIAYPRDD
LRGRVDCSRAGFLQAMDISWSPFIRMAHPLMKGTITFCMTTYGSGVMVYHNM
MGPVKAALSESATRYLAELGPGQIRVHAISPCLATRAASGITGFDALLDRAQAKAPA
RSLVGIDVDATAMLATDAARATGTTTGGCCCTGTGGCATCTGACGTACTGCGCGTGATCC
3070
185368
3071
185308
3131
185248
3191
185191
3251
185131
3311
185071
3371
185011
3431
184951
3491
```





Db 317524 GCGCGATCCGGGCTGGCGCGGTGACCAAGGCGGTCTGCGGCGCAAGG 317574

Search completed: January 18, 2006, 16:18:34  
Job time : 21080 secs